

# Could we build a conscious robot?

Owen Holland



University of Essex

*Department of*  
**Computer Science**

What is machine consciousness?

# What is machine consciousness?

A new area of research dedicated to the construction of machines that are conscious like you...

# What is machine consciousness?

A new area of research dedicated to the construction of machines that are conscious like you...

...really conscious, not just mimicking consciousness...



# What is machine consciousness?

A new area of research dedicated to the construction of machines that are conscious like you...

...really conscious, not just mimicking consciousness...

...and with real feelings, not just simulated feelings

# What is machine consciousness?

A new area of research dedicated to the construction of machines that are conscious like you...

...really conscious, not just mimicking consciousness...

...and with real feelings, not just simulated feelings

This is **STRONG** machine consciousness.

# What is machine consciousness?

A new area of research dedicated to the construction of machines that are conscious like you...

...really conscious, not just mimicking consciousness...

...and with real feelings, not just simulated feelings

This is STRONG machine consciousness.

WEAK machine consciousness is aimed at machines that ***behave*** as if they were conscious.

## It has a past...

2001: Can a machine be conscious? (Swartz Foundation, Cold Spring Harbor Laboratory)

2003: Computational models of consciousness (ASSC7, Memphis)

Models of consciousness (European Science Foundation, Birmingham)

Machine consciousness: Complexity aspects (EU Complex Systems Network of Excellence, Turin)

2004: Machine models of consciousness workshop (Antwerp)

2005: Next generation approaches to machine consciousness (AISB05, U of Hertfordshire)

International workshop on artificial consciousness (Accademia di Studi Mediterranei, Agrigento)

...and a future

2006: Integrative approaches to machine consciousness  
(AISB06)

International Symposium on Machine Models of  
Consciousness (ICSC, Lesvos, Greece)

# What is consciousness?

...you've got to distinguish between the scientific definition that comes at the end of the investigation where we now know how it works, and the common-sense definition that you start off with, the aim of which is to identify the target.

John Searle 2004

## What is consciousness?

Consciousness is defined as those states of sentience or feeling or awareness that begin in the morning when you wake up from a dreamless sleep, and they continue on all day long until you fall asleep again, get hit over the head and knocked unconscious, or go into a coma, or die, or otherwise, as we would say, become unconscious...That's the target.

John Searle 2004

## What is consciousness?

‘I have assumed that consciousness exists, and that to redefine the problem as that of explaining how certain cognitive or behavioural functions are performed is unacceptable...*If you hold that an answer to the ‘easy’ problems explains everything that needs to be explained, then you get one sort of theory; if you hold that there is a further ‘hard’ problem, then you get another.*’

David Chalmers ‘The Conscious Mind’ 1996



## What is consciousness?

‘...I must present a theory (of sentience) that addresses questions like these: If we could ever duplicate the information processing in the human mind as an enormous computer program, would a computer running the program be conscious?... etc...etc...’

Steven Pinker ‘How the Mind Works’ 1997

## What is consciousness?

‘...I must present a theory (of sentience) that addresses questions like these: If we could ever duplicate the information processing in the human mind as an enormous computer program, would a computer running the program be conscious?... etc...etc...’

Beats the heck out of me! I have some prejudices, but no idea of how to look for a defensible answer. And neither does anyone else.’

Steven Pinker ‘How the Mind Works’ 1997

Common sense tells us it's obvious...

- that we consciously perceive the world accurately
- that we consciously remember what we perceive
- that we consciously decide on actions, and then consciously initiate and control them
- etc etc

Common sense tells us it's obvious...

- that we consciously perceive the world accurately
- that we consciously remember what we perceive
- that we consciously decide on actions, and then consciously initiate and control them
- etc etc

We teach our children this

The law assumes we do this

etc etc

...but common sense is wrong

Change and inattention blindness (Simons, O'Regan, etc)

- you don't see what's there

Misattribution of agency (Daprati, Wegner)

- you don't know your own actions

Backwards referral of sensation (Libet)

- when sensations become conscious, they are experienced as if they started about half a second previously

Backwards referral of action (Walter, Kornhuber)

- the neural processes of a voluntary action begin about half a second before you are aware of initiating it

‘Consciousness is a peculiar phenomenon. It is riddled with deceit and self-deception; there can be consciousness of something we were sure had been erased by an anaesthetic; the conscious / is happy to lie up hill and down dale to achieve a rational explanation for what the body is up to; sensory perception is the result of a devious relocation of sensory input in time; when the consciousness thinks it determines to act, the brain is already working on it; there appears to be more than one version of consciousness present in the brain; our conscious awareness contains almost no information but is perceived as if it were vastly rich in information. Consciousness is peculiar.’

Tor Norretranders ‘The User Illusion’ 1991 (tr 1998)

So:

we can't really define it

we don't know how it arises

and it's peculiar

So:

we can't really define it  
we don't know how it arises  
and it's peculiar

And that's what makes it interesting.



So:

we can't really define it  
we don't know how it arises  
and it's peculiar

And that's what makes it interesting.

Is some critical problem being solved by the  
apparently ramshackle nature of phenomenal  
consciousness?

So:

we can't really define it  
we don't know how it arises  
and it's peculiar

And that's what makes it interesting.

Is some critical problem being solved by the  
apparently ramshackle nature of phenomenal  
consciousness?

Or is it a side effect of the solution of some critical  
problem?

So:

we can't really define it  
we don't know how it arises  
and it's peculiar

And that's what makes it interesting.

Is some critical problem being solved by the  
apparently ramshackle nature of phenomenal  
consciousness?

Or is it a side effect of the solution of some critical  
problem?

We just don't know – yet.

So how could we build a conscious machine?

So how could we build a conscious machine?

(1) Identify the components of consciousness, and implement all of them in a machine?

Identify the components of consciousness, and implement all of them in a machine?

Bernard Baars' theory of the global workspace is currently the most detailed and well supported theory of consciousness. It identifies components and mechanisms.

Stan Franklin (University of Memphis) is working with Baars to produce 'conscious software', with each component and mechanism of Baars' system represented in a complex piece of software.

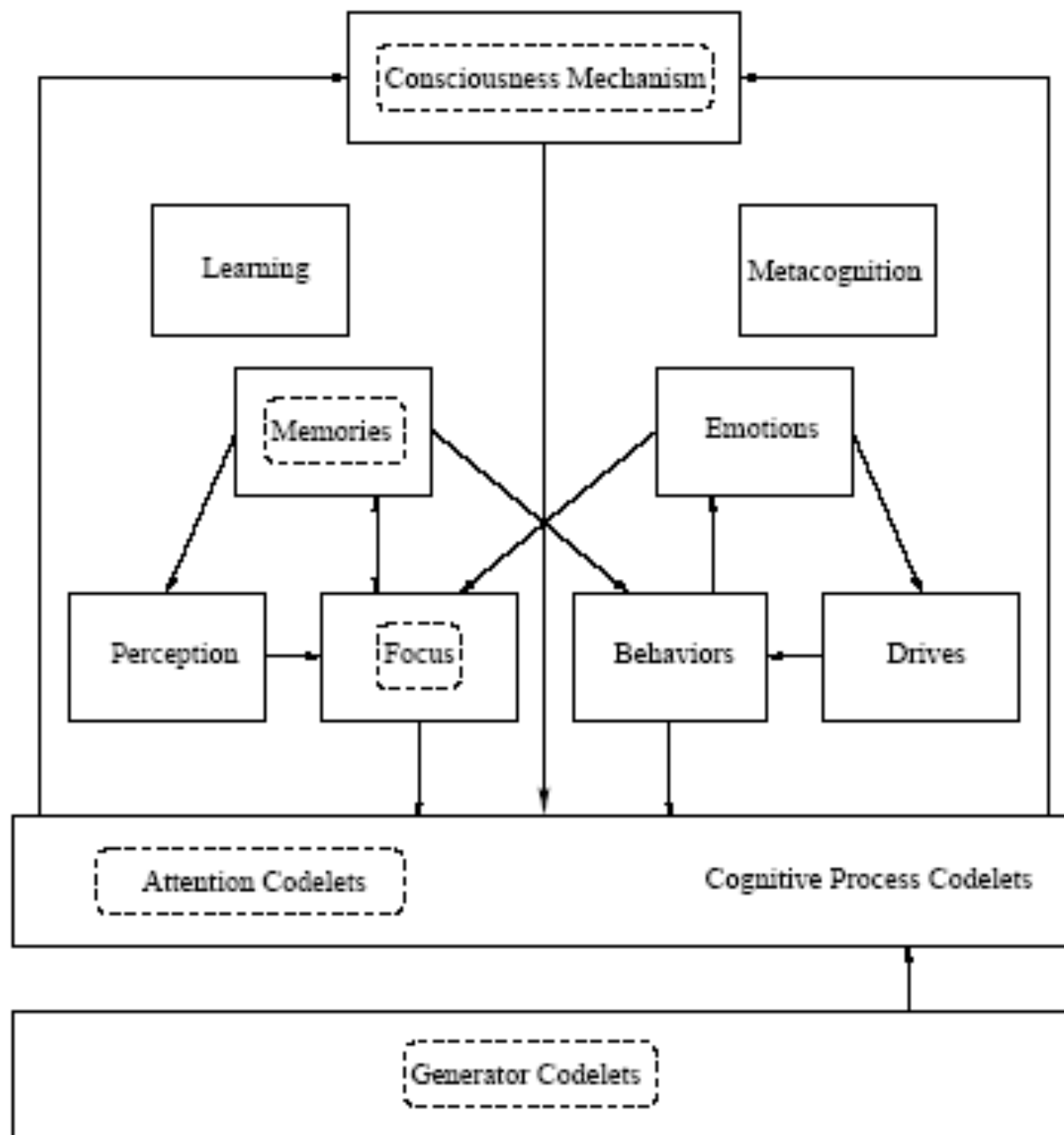


Fig. 2. An Architecture For “Conscious” Software Agents

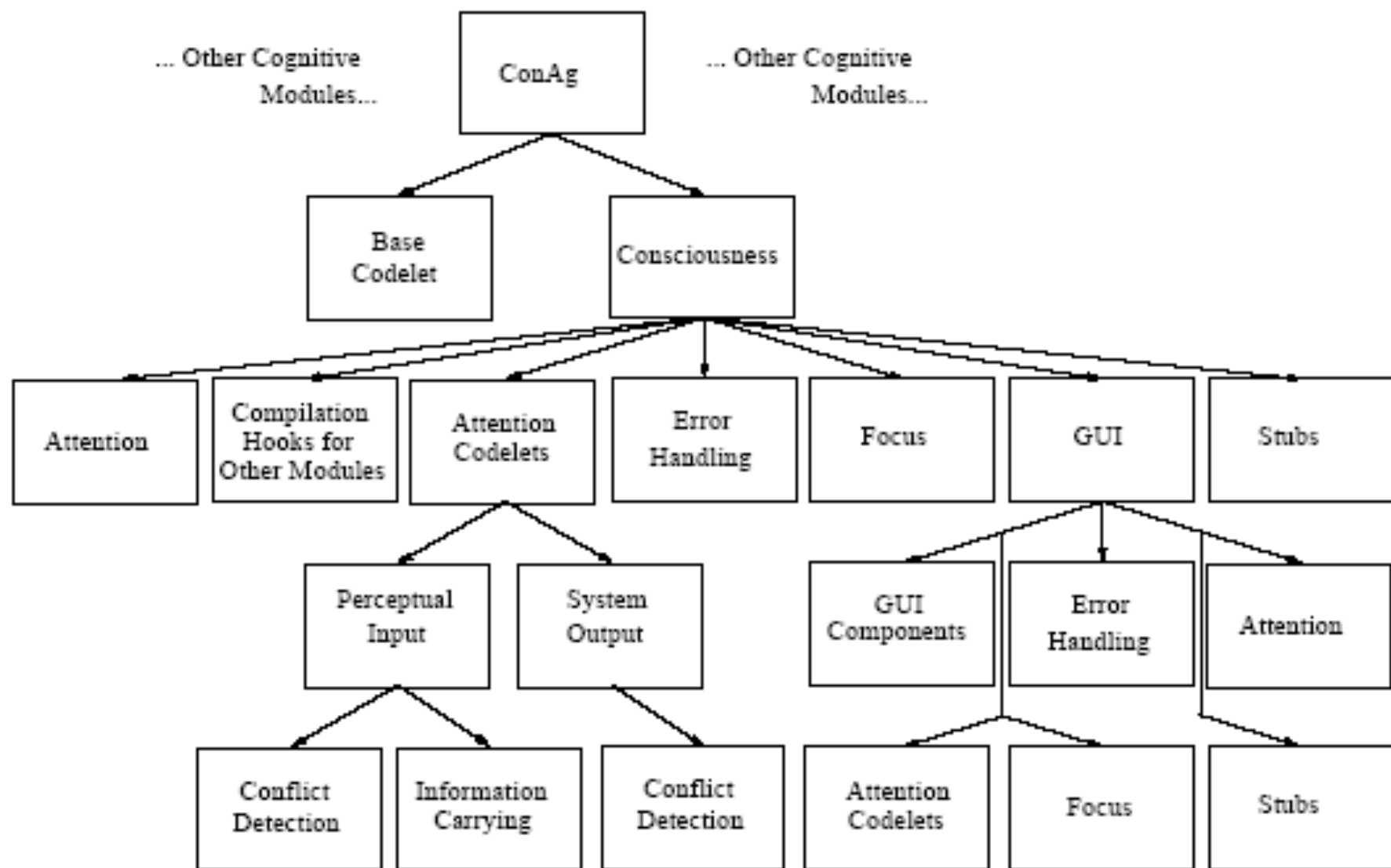


Fig. 3. ConAg's Hierarchical Framework



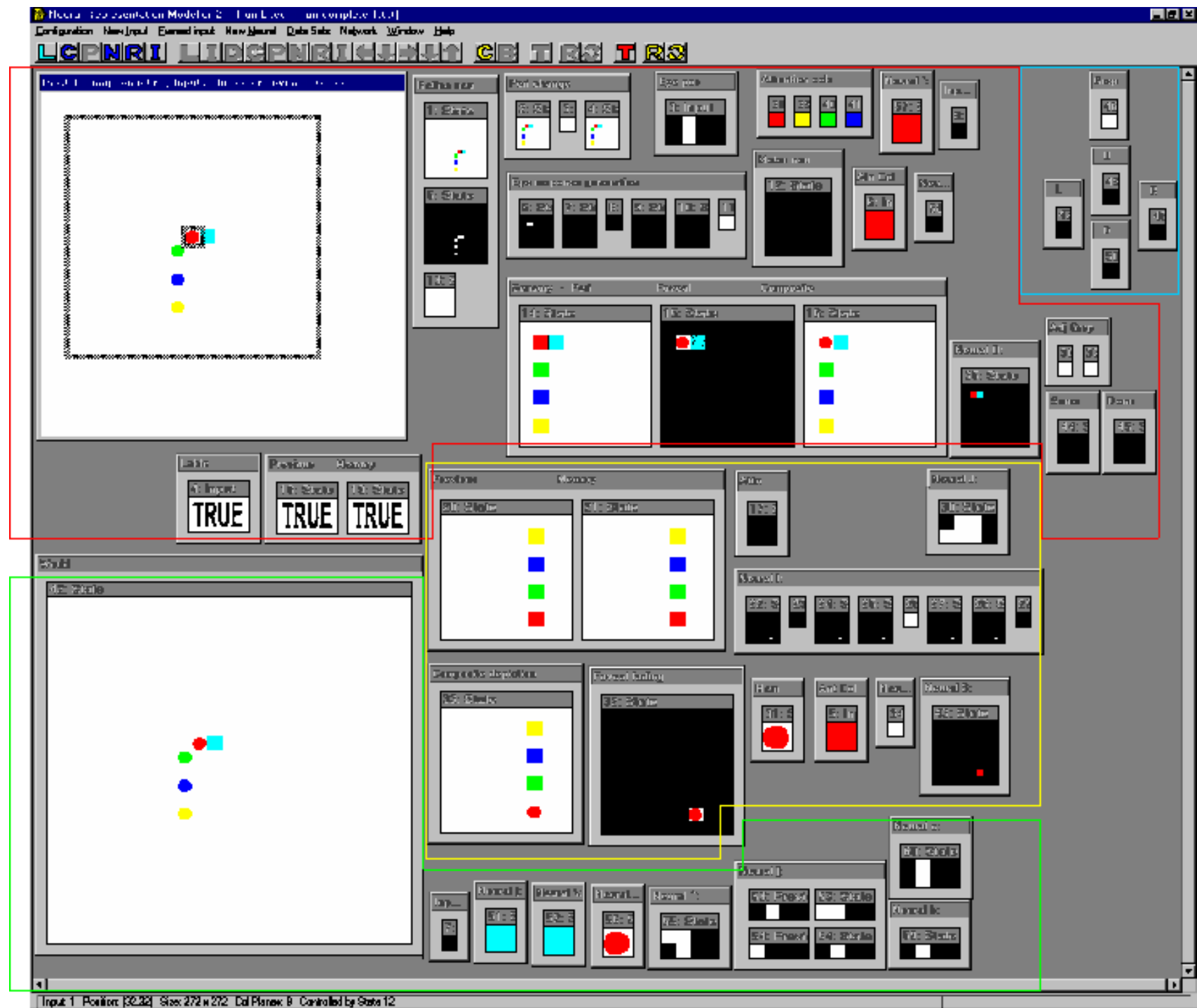
So how could we build a conscious machine?

(1) Identify the components of consciousness, and implement all of them in a machine?

(2) Identify the components of the machine that produces consciousness (the brain) and copy them?

Identify the components of the machine that produces consciousness (the brain) and copy them?

This is probably the most popular approach. An example is the work of Igor Aleksander, who models the process of visual consciousness by building and interconnecting neural network models of the structures known to be involved.



So how could we build a conscious machine?

(1) Identify the components of consciousness, and implement all of them in a machine?

(2) Identify the components of the machine that produces consciousness (the brain) and copy them?

(3) Identify the circumstances in which consciousness arose, copy them, and hope that consciousness emerges again?

So how could we build a conscious machine?

(1) Identify the components of consciousness, and implement all of them in a machine?

(2) Identify the components of the machine that produces consciousness (the brain) and copy them?

(3) ***Identify the circumstances in which consciousness arose, copy them, and hope that consciousness emerges again?***

How did consciousness arise?

How did consciousness arise?

We don't know (just as we don't know exactly what consciousness *is*) but it was probably something to do with the development of high intelligence

How did consciousness arise?

We don't know (just as we don't know exactly what consciousness *is*) but it was probably something to do with the development of high intelligence

How did intelligence arise?



How did consciousness arise?

We don't know (just as we don't know exactly what consciousness *is*) but it was probably something to do with the development of high intelligence

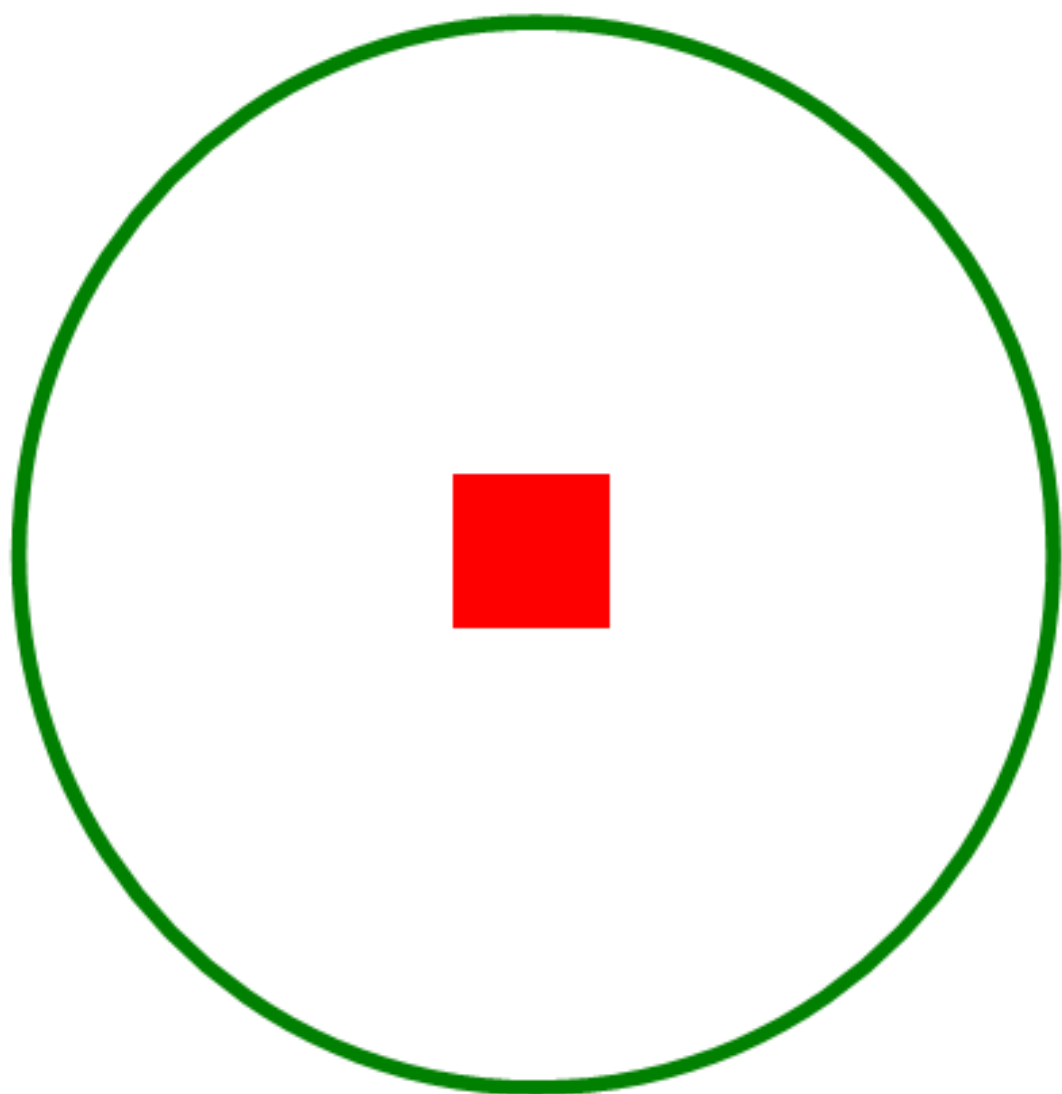
How did intelligence arise?

Through natural and sexual selection – and we can almost understand how and why

Let's consider the problems of an autonomous embodied agent (an animal or robot)...



Let's consider the problems of an autonomous embodied agent (an animal or robot) in a complex, occasionally novel, dynamic, and hostile world...



Let's consider the problems of an autonomous embodied agent (an animal or robot) in a complex, occasionally novel, dynamic, and hostile world, in which it has to achieve some task (or mission).

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency?

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No



How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No
- by having learned the consequences for the achievement of the mission of every possible action in every contingency?

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No
- by having learned the consequences for the achievement of the mission of every possible action in every contingency? No

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No

- by having learned the consequences for the achievement of the mission of every possible action in every contingency? No

- by having **learned enough** to be able to **predict the consequences** of tried and untried actions, by being able to **evaluate** those consequences for their likely **contribution to the mission**, and by **selecting** a relatively **good** course of action?

How could the agent achieve its task (or mission)?

- by being preprogrammed for every possible contingency? No

- by having learned the consequences for the achievement of the mission of every possible action in every contingency? No

- by having **learned enough** to be able to **predict the consequences** of tried and untried actions, by being able to **evaluate** those consequences for their likely **contribution to the mission**, and by **selecting** a relatively **good** course of action? Maybe...

But how could it predict?

But how could it predict?

For actions it has tried before in these circumstances, it could simply remember what happened last time

But how could it predict?

For actions it has tried before in these circumstances, it could simply remember what happened last time

If things are only slightly different, it could simply generalise from what it has learned

But how could it predict?

For actions it has tried before in these circumstances, it could simply remember what happened last time

If things are only slightly different, it could simply generalise from what it has learned

Otherwise, it could run some kind of ***simulation*** of its potential actions in the world, enabling it to predict their effects – ***even if they involve novel situations or actions***



Here's how Richard Dawkins puts it:

“Survival machines that can simulate the future are one jump ahead of survival machines who can only learn on the basis of overt trial and error.”

Dawkins, 1976, *The Selfish Gene*

Two questions:

What exactly has to be simulated?

What is needed for simulation?

What exactly has to be simulated?

Whatever affects the mission. In an ***embodied*** agent, the agent can only affect the world through the actions of its body in and on the world, and the world can only affect the mission by affecting the agent's body.

What exactly has to be simulated?

Whatever affects the mission. In an ***embodied*** agent, the agent can only affect the world through the actions of its body in and on the world, and the world can only affect the mission by affecting the agent's body.

So it needs to simulate those aspects of its **BODY** that affect the world in ways that affect the mission, along with those aspects of the **WORLD** that affect the body in ways that affect the mission.

## What is needed for simulation?

Some structures or processes corresponding to states of the world that, when operated on by processes or structures corresponding to actions, yields outcomes corresponding to the consequences of those actions.

## What is needed for simulation?

Some structures or processes corresponding to states of the world that, when operated on by processes or structures corresponding to actions, yields outcomes corresponding to the consequences of those actions.

I like to call these structures or processes 'internal models', because they are like working models rather than images or static representations

What is needed for simulation?

So we require a model (or linked set of models) that includes the body, and how it is controlled, and the spatial aspects of the world, and the (kinds of) objects in the world, and their spatial arrangement. But consider...

What is needed for simulation?

The **body** is always present and available, and changes slowly, if at all. When it moves, it is usually because it has been commanded to move.



## What is needed for simulation?

The **body** is always present and available, and changes slowly, if at all. When it moves, it is usually because it has been commanded to move.

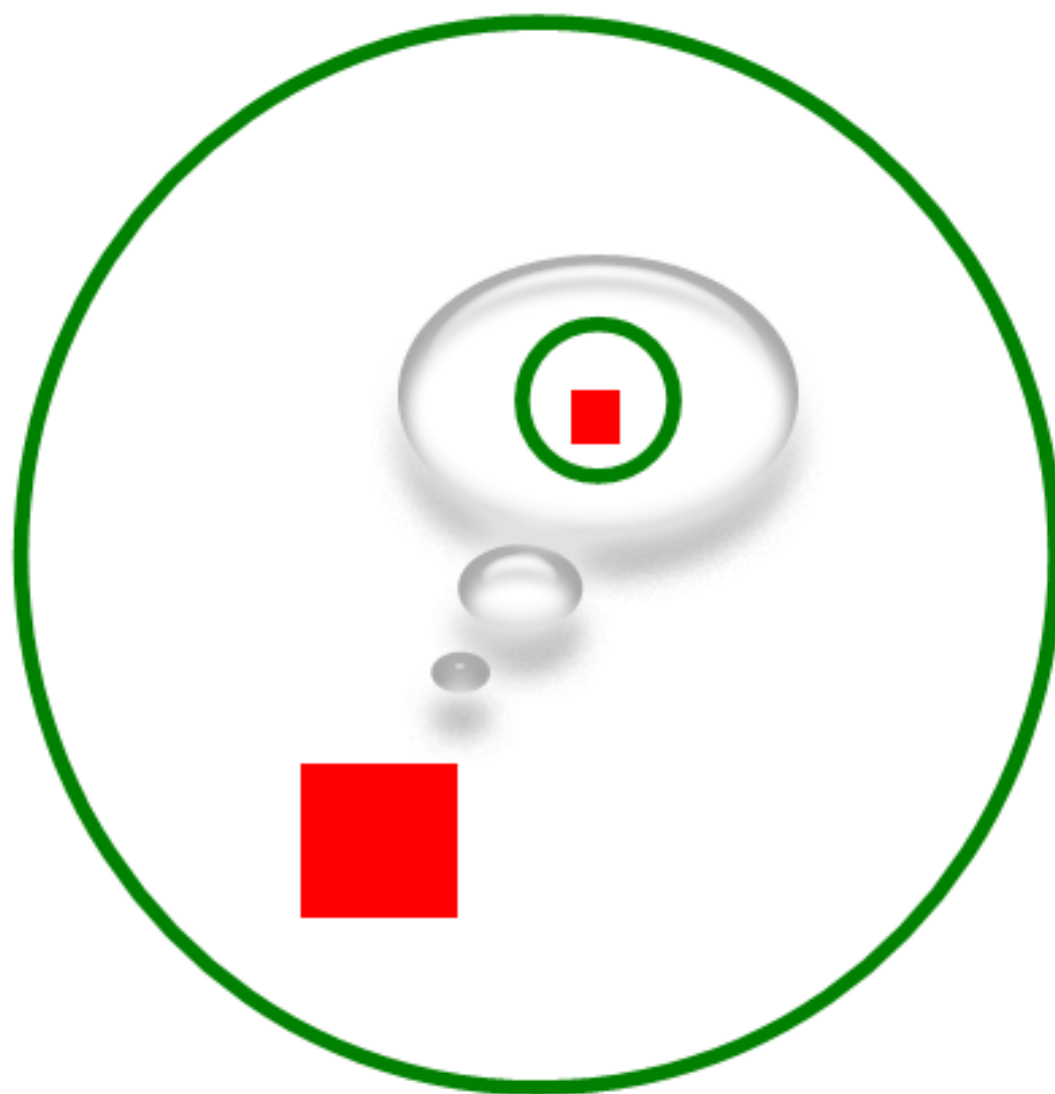
The **world** is different. It is 'complex, occasionally novel, dynamic, and hostile'. It's only locally available, and may contain objects of known and unknown kinds in known and unknown places.

What is needed for simulation?

The **body** is always present and available, and changes slowly, if at all. When it moves, it is usually because it has been commanded to move.

The **world** is different. It is 'complex, occasionally novel, dynamic, and hostile'. It's only locally available, and may contain objects of known and unknown kinds in known and unknown places.

How should all this be modelled? As a single model containing both body and world?

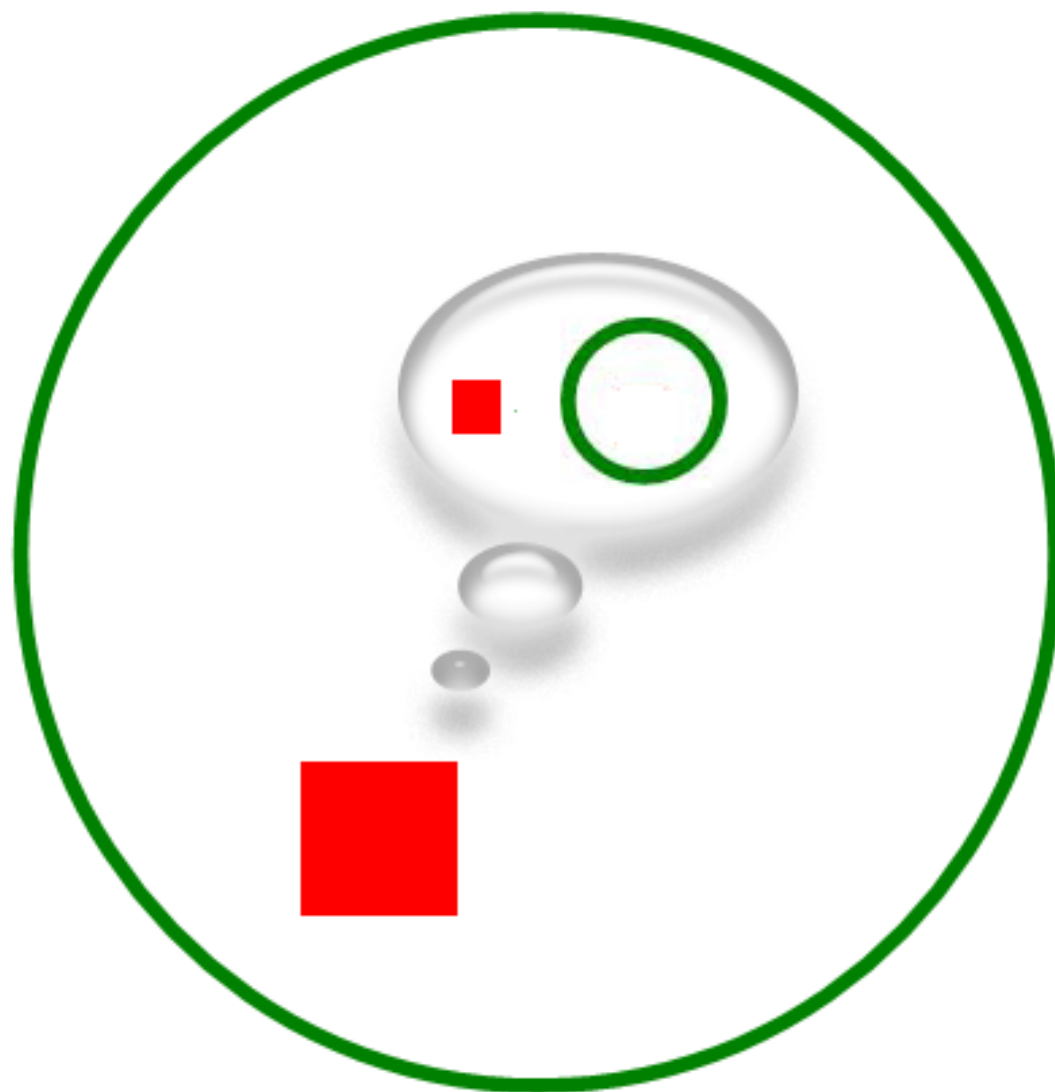


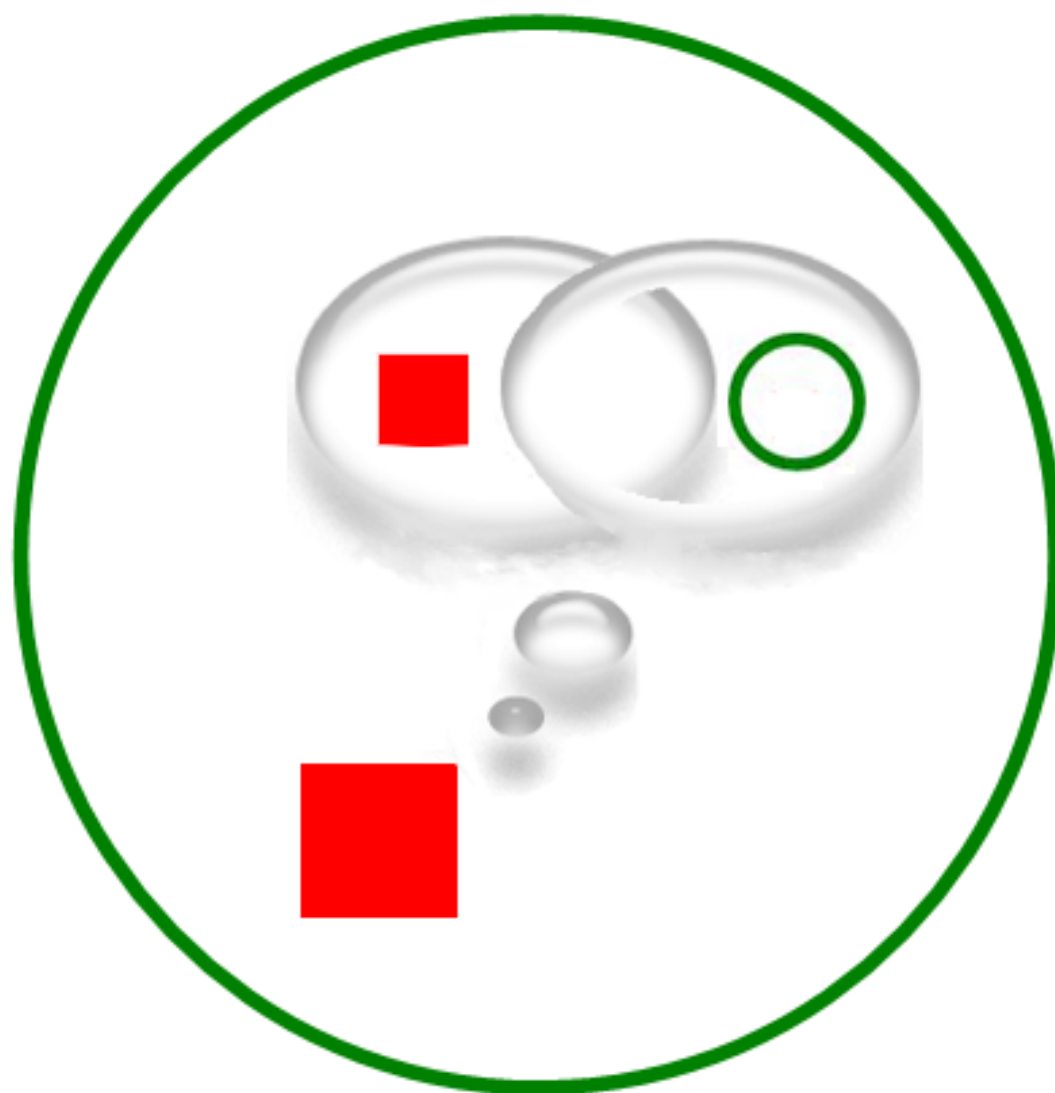
What is needed for simulation?

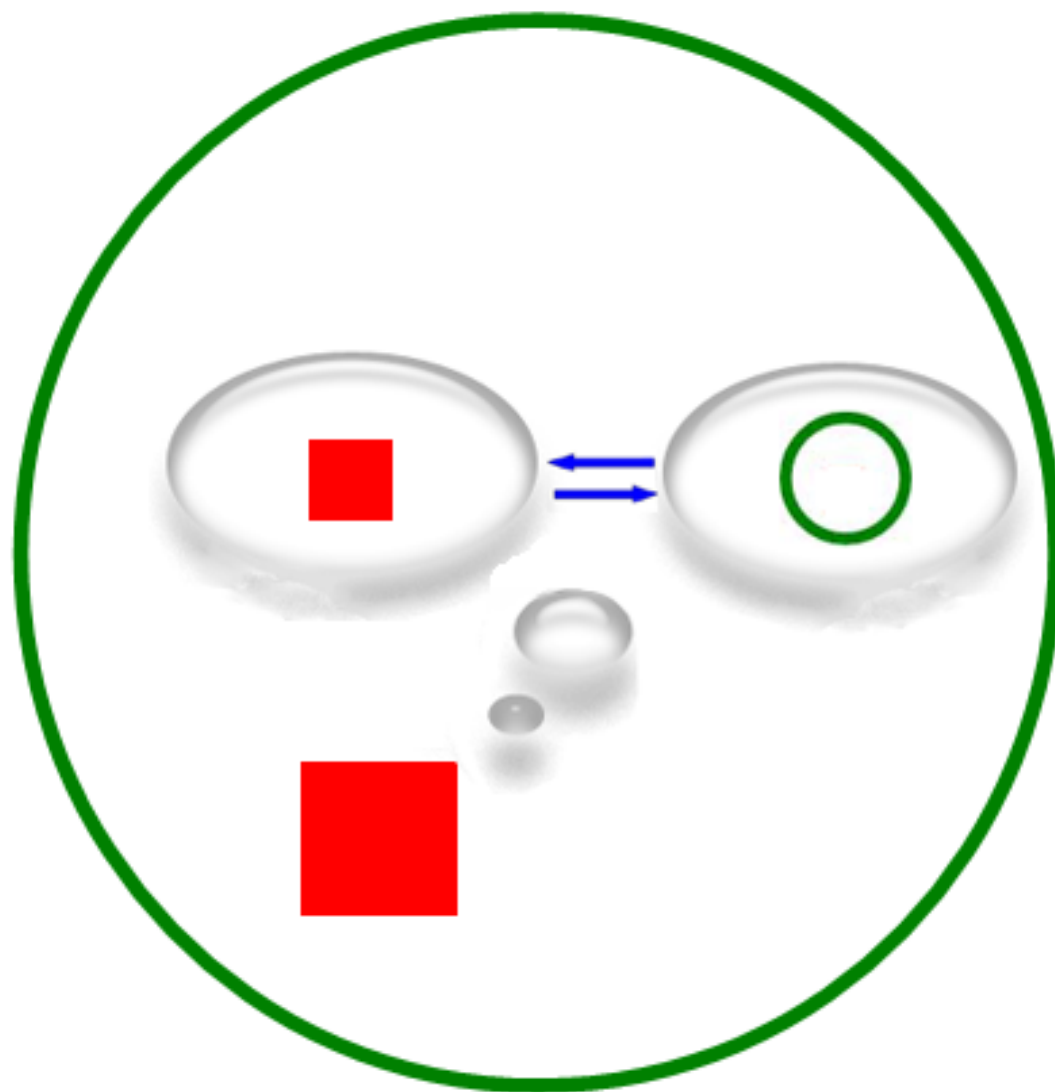
The **body** is always present and available, and changes slowly, if at all. When it moves, it is usually because it has been commanded to move.

The **world** is different. It is 'complex, occasionally novel, dynamic, and hostile'. It's only locally available, and may contain objects of known and unknown kinds in known and unknown places.

How should all this be modelled? As a single model containing both body and world? ***Or as a separate model of the body coupled to and interacting with a separate model of the world?***







“...(I)t is always obvious to you that there are some things you can do and others you cannot given the constraints of your body and of the external world. (You know you can't lift a truck...) Somewhere in your brain there are representations of all these possibilities, and the systems that plan commands...need to be aware of this distinction between things they can and cannot command you to do....***To achieve all this, I need to have in my brain not only a representation of the world and various objects in it but also a representation of myself, including my own body within that representation....In addition, the representation of the external object has to interact with my self-representation....***”

(Ramachandran and Blakeslee 1998).



Does the brain model the body?

Yes, in many ways. It models the muscular control of movement. It also predicts the nature and timing of the internal and external sensory inputs that will be produced if the movement is executed correctly.

## Does the brain model the body?

Yes, in many ways. It models the muscular control of movement. It also predicts the nature and timing of the internal and external sensory inputs that will be produced if the movement is executed correctly.

Ramachandran and Blakeslee describe a host of body image phenomena involving phantom limbs. In one case, a patient with congenital absence of both arms had apparently 'normal' phantom limbs from an early age. Some components of the internal model of the body may be inborn.

## Does the brain model the body?

Yes, in many ways. It models the muscular control of movement. It also predicts the nature and timing of the internal and external sensory inputs that will be produced if the movement is executed correctly.

Ramachandran and Blakeslee describe a host of body image phenomena involving phantom limbs. In one case, a patient with congenital absence of both arms had apparently 'normal' phantom limbs from an early age. Some components of the internal model of the body may be inborn.

But some people have very unusual body plans, yet manage perfectly well...

Myrtle  
Corbin  
b 1868



Myrtle  
Corbin  
b 1868



# The Commerce Journal.

---

COMMERCE, HUNT COUNTY, TEXAS, FRIDAY, JULY 22, 1910.

---

## WOMAN WITH FOUR LEGS.

---

A World Wonder Now Visiting  
Relatives Near Commerce,  
Was At Picnic.

---

A most unusual woman has been visiting relatives near this city for several weeks and, with her husband, attending picnics, at which places she is on exhibition in a tent. In brief, she

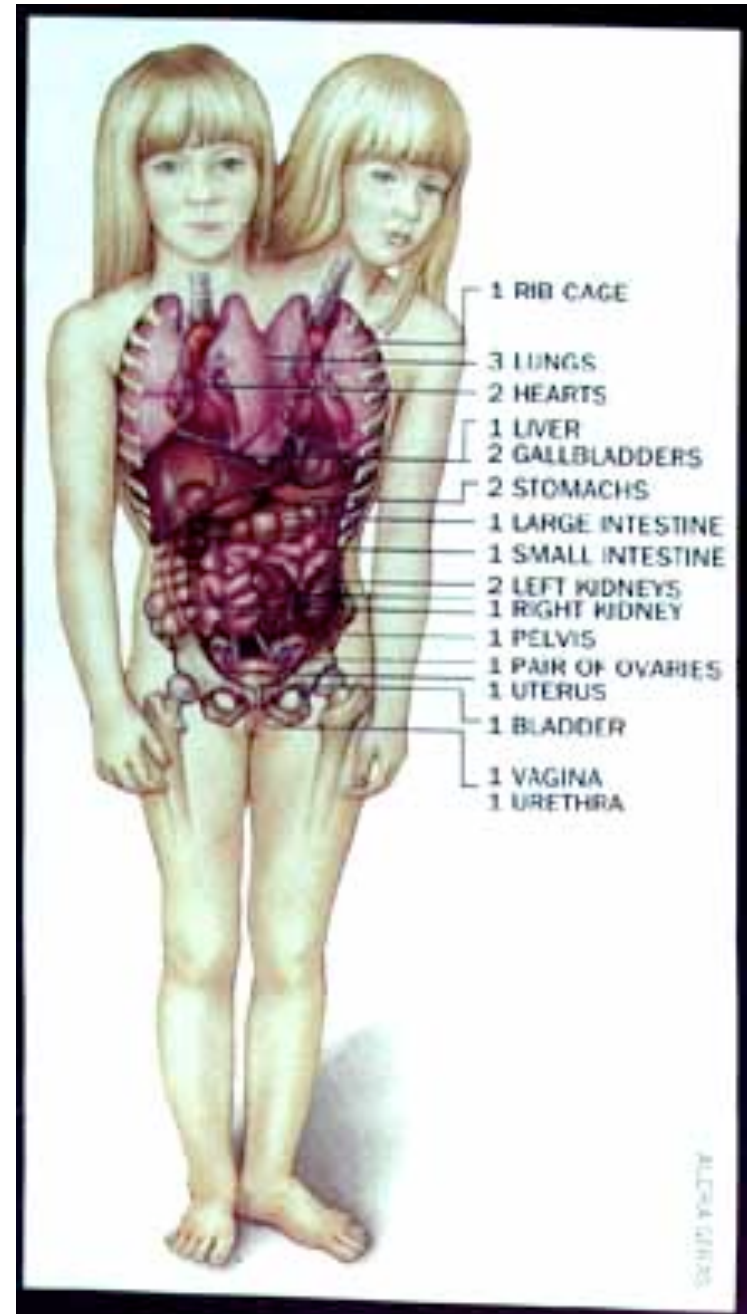
One of the surprising things about Mrs. Bicknell is that while she has four limbs, as shown in the accompanying picture, the two inside ones are smaller than the others, thus enabling her to get about with no inconvenience and without any one suspecting her true condition. Her youngest child is a pretty, red cheeked baby of some two years, and like its brothers and sisters, is perfectly developed.

---

# Abigail and Brittany Hensel



# Abigail and Brittany Hensel





Does the brain model the world?

Yes, in many ways. It models space, and it models the nature and behaviour of objects, and much of this modelling is innate.

Useful reading (for me anyway):

Wild Minds, by Marc Hauser.

Folk Physics for Apes, by Daniel Povinelli

What has this to do with consciousness?

What Dawkins (1976) said next:

“Survival machines that can simulate the future are one jump ahead of survival machines who can only learn on the basis of overt trial and error...***The evolution of the capacity to simulate seems to have culminated in subjective consciousness... Perhaps consciousness arises when the brain’s simulation of the world becomes so complete that it must include a model of itself.***”

In other words...

Intelligence may depend on the possession and manipulation of an internal model of the agent (the IAM) interacting with an internal model of the world

**AND**

the presence and interaction of these models may also underlie the production of consciousness.

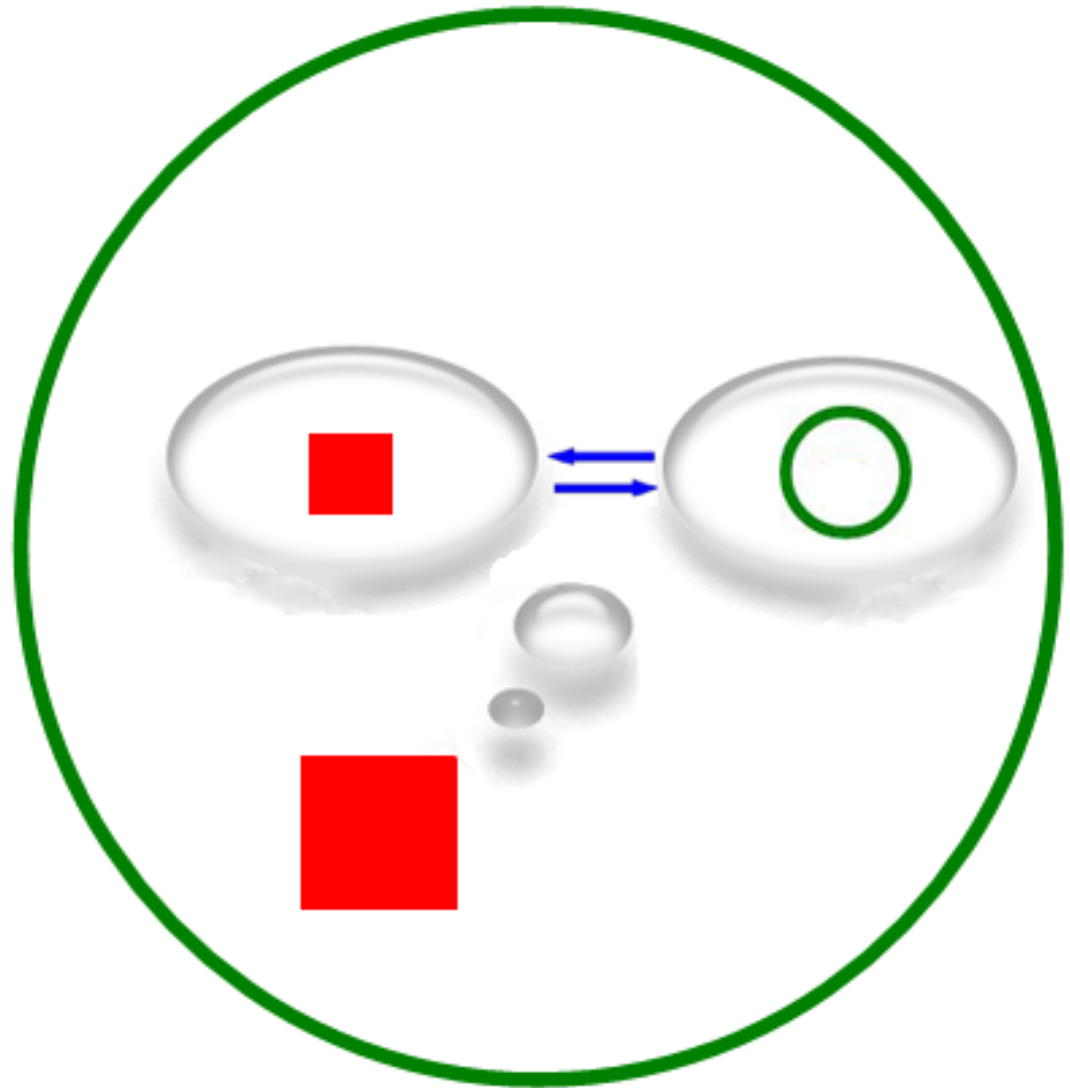
"...consciousness requires that the brain must represent not just the object, not just a basic self structure, ***but the interaction of the two***....This is still an atypical foundation for a theory of consciousness, given that until recently, it was implicitly assumed that the self could be left out of the equation. There has been a recent sea change on this crucial point..."

Douglas Watt 2000, review of Damasio's "The Feeling of What Happens" (Damasio 1999).

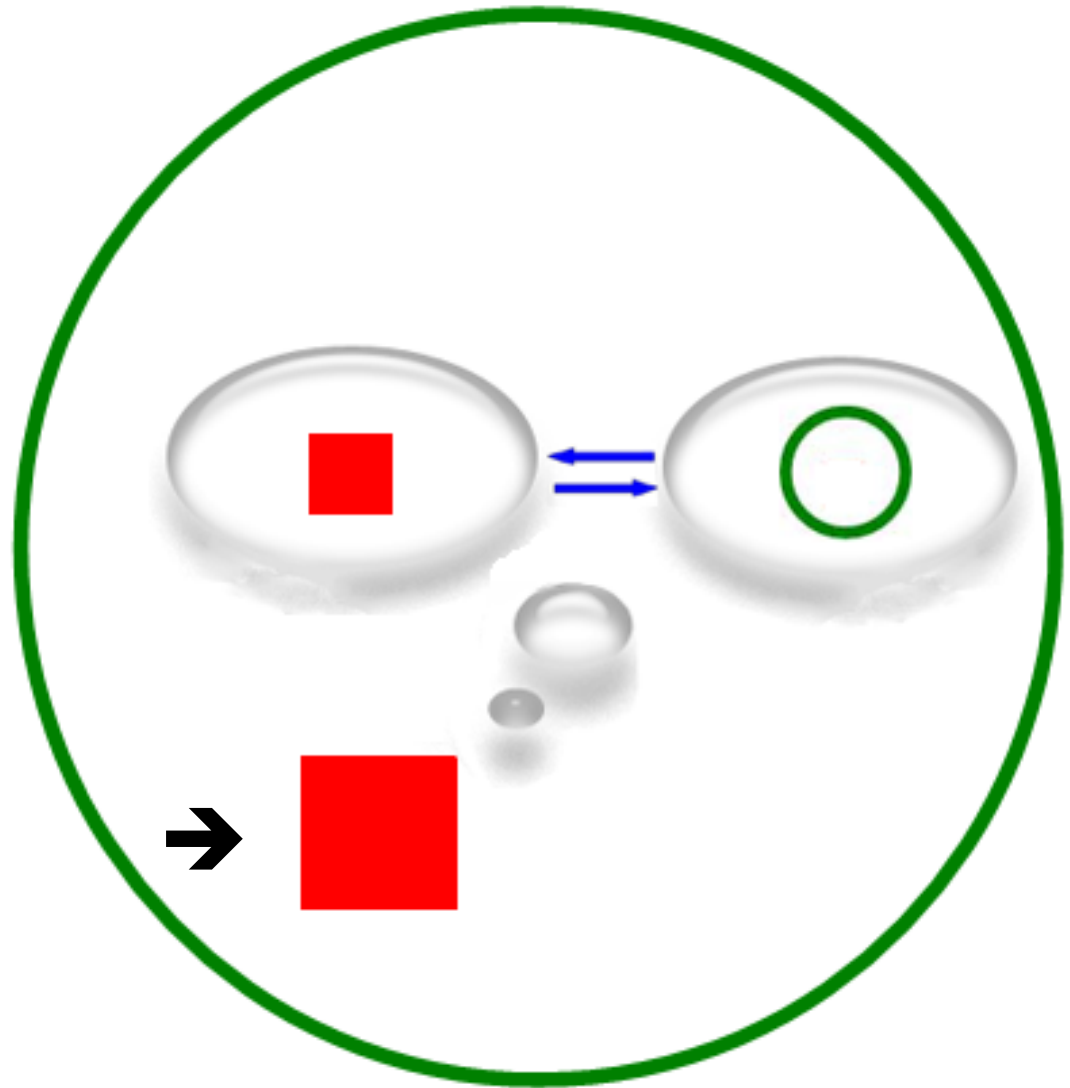
Back to our  
agent...

Consider things  
from the point of  
view of the internal  
agent model  
(IAM).

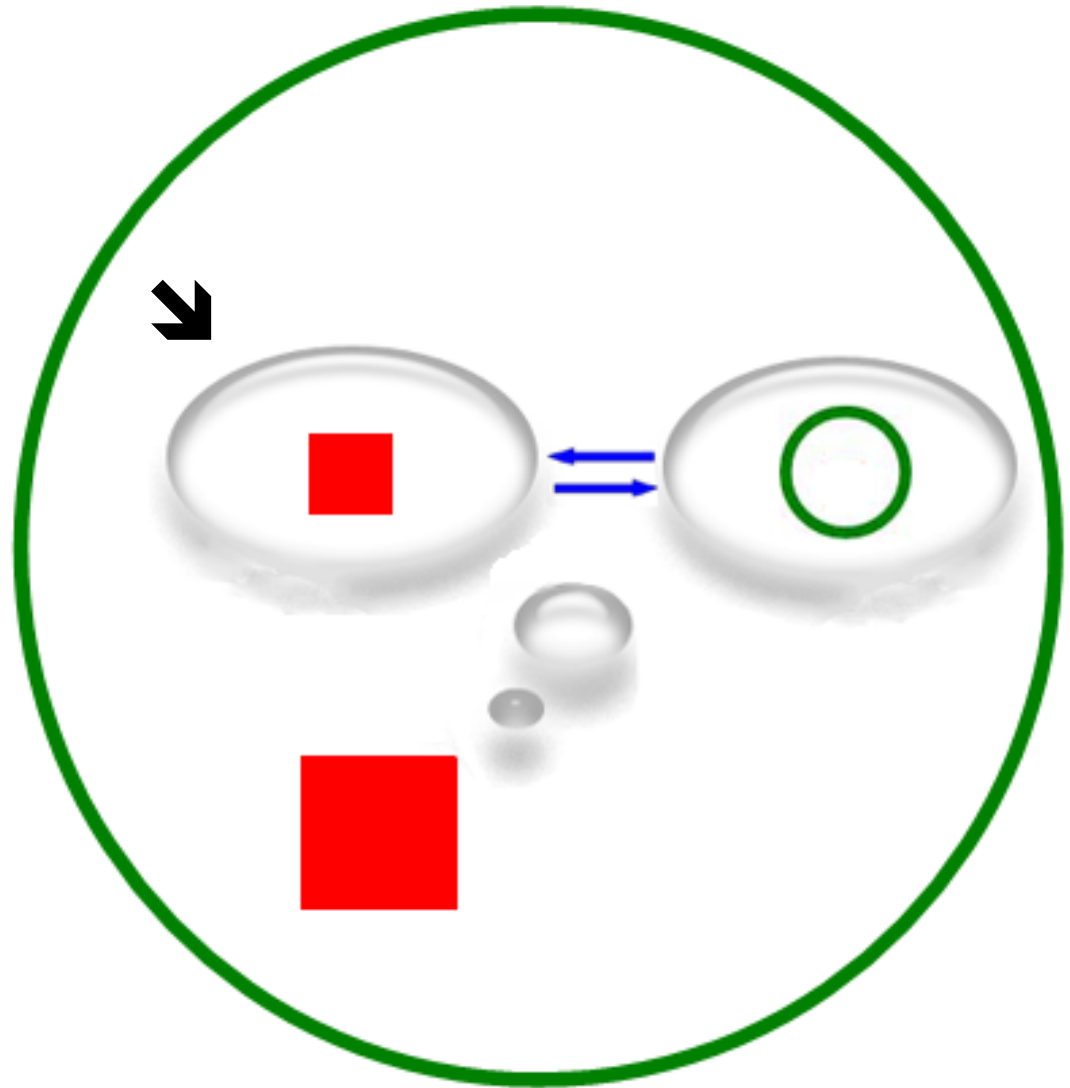
Suppose the  
simulations are  
really good...



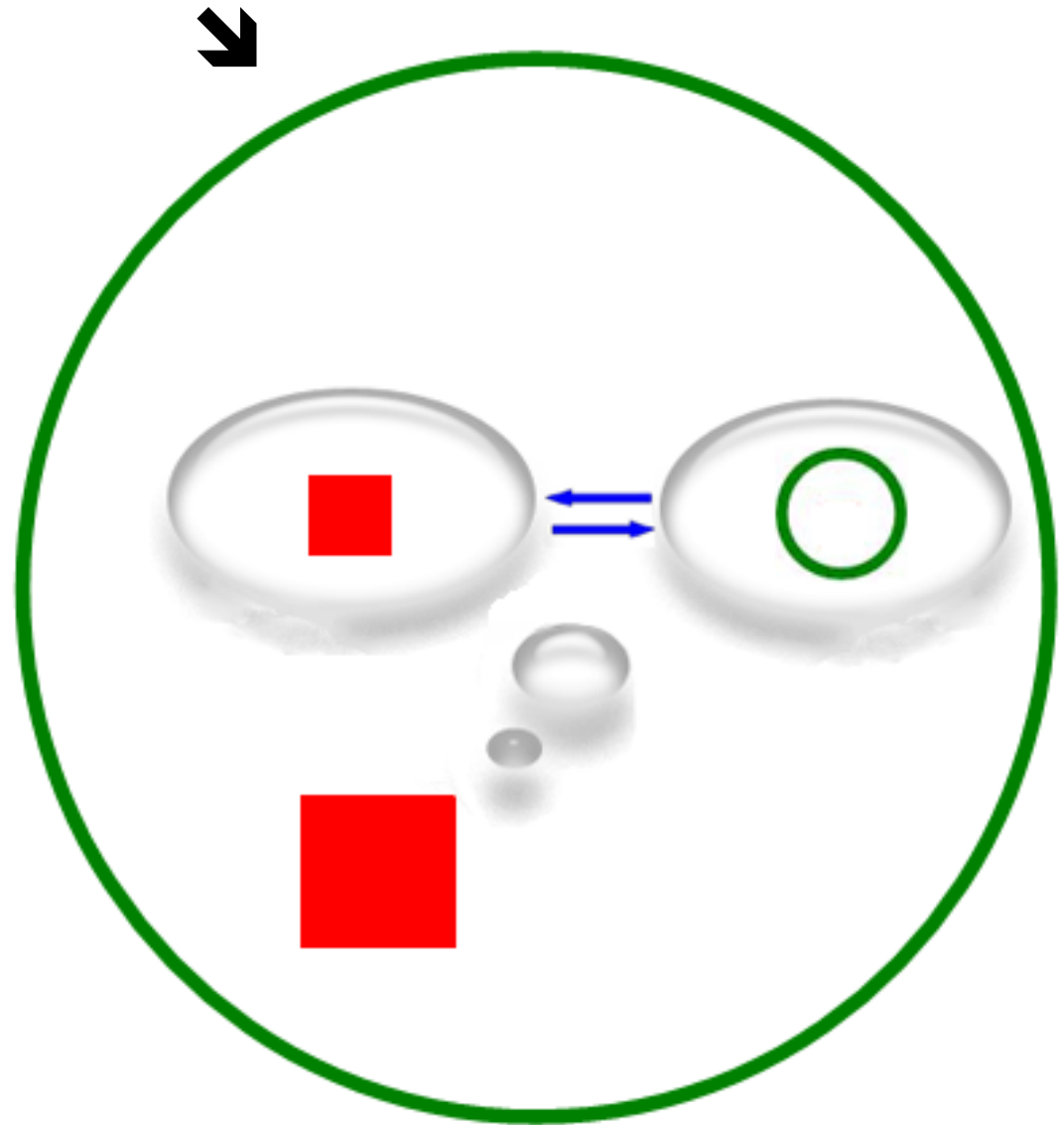
It (the model)  
will 'think' it's  
this, the  
agent...



...but it's actually  
this – a model of  
the agent

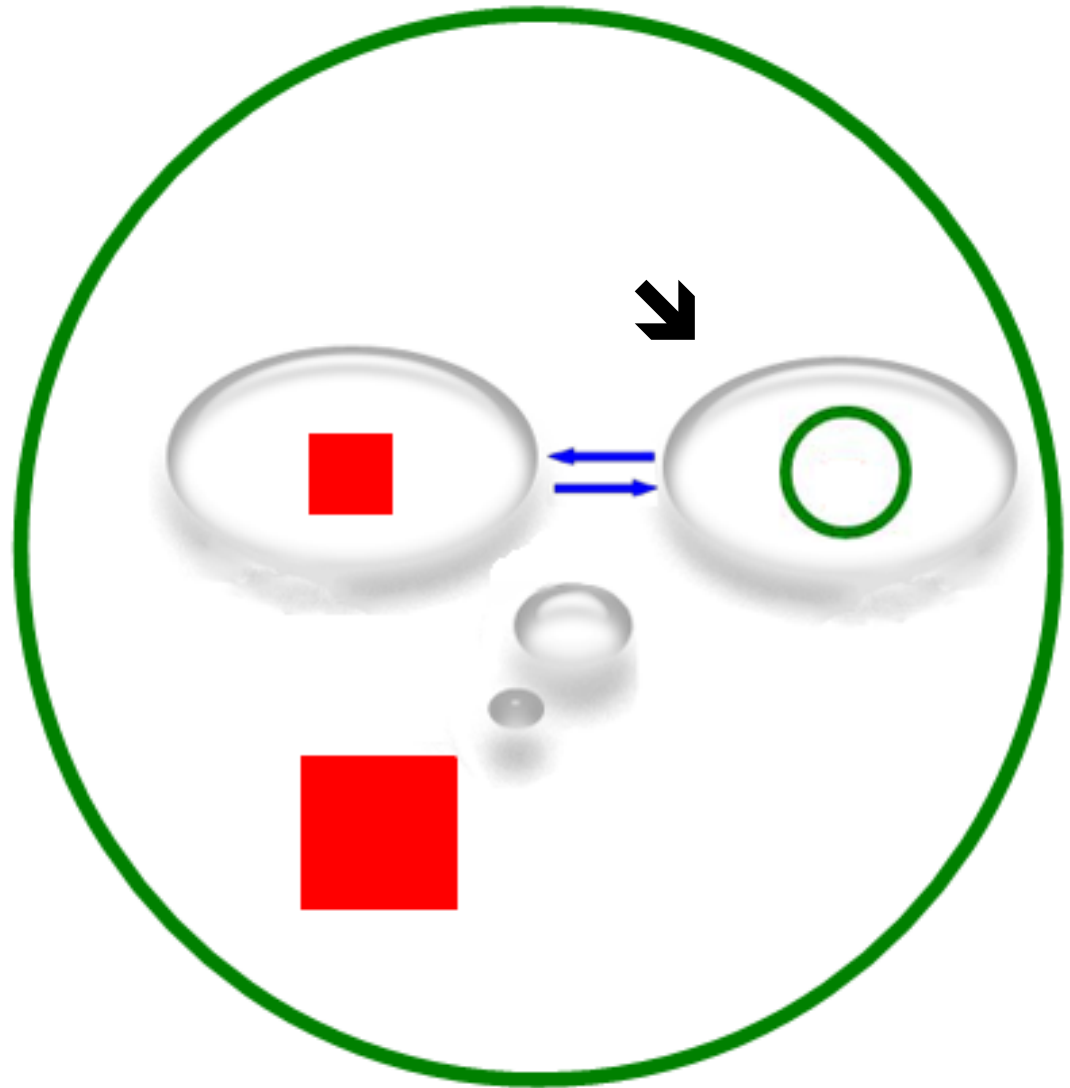


It will 'think' it's  
interacting with the  
real world





But it's actually  
interacting with a  
model of the real  
world



# Blackmore's hypothesis

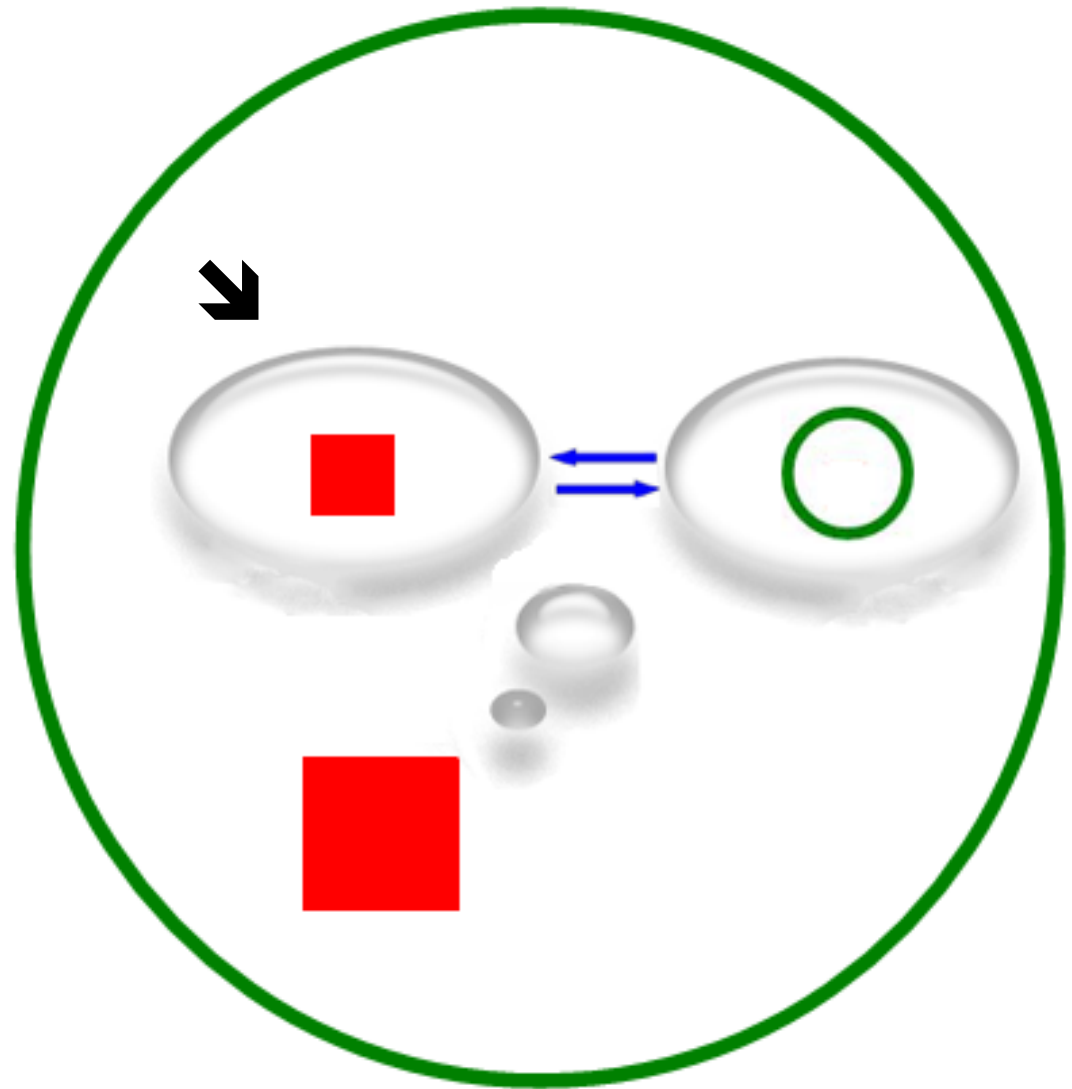
In the late 1980s, the psychologist Sue Blackmore proposed that the conscious self – the experiencing entity – was in fact an internal self model.

# Blackmore's hypothesis

In the late 1980s, the psychologist Sue Blackmore proposed that the conscious self – the experiencing entity – was in fact an internal self model.

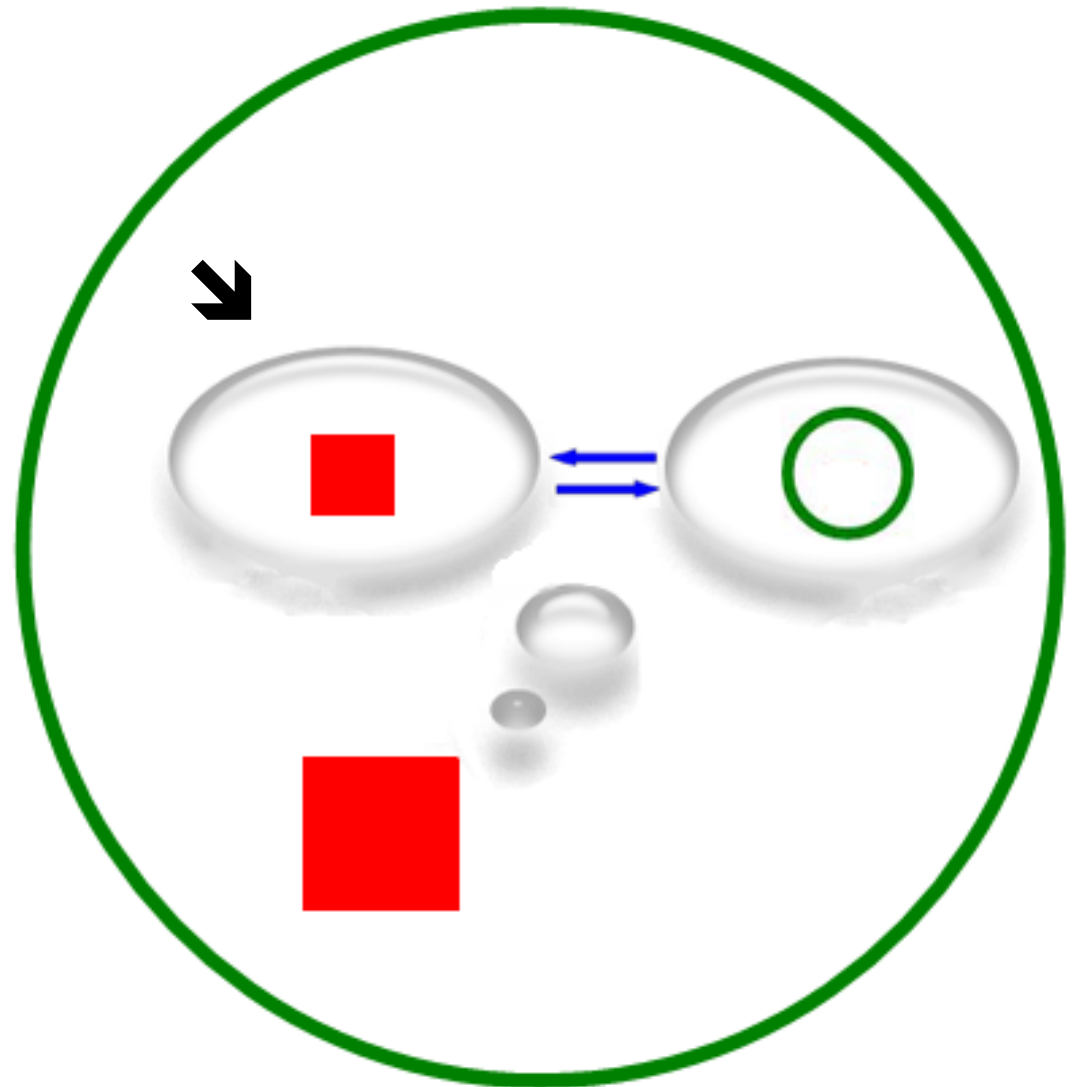
No-one took the idea seriously at the time – but if it is true...

Consciousness and feelings are in the Internal Agent Model (the IAM) – the system's 'software' model of itself...

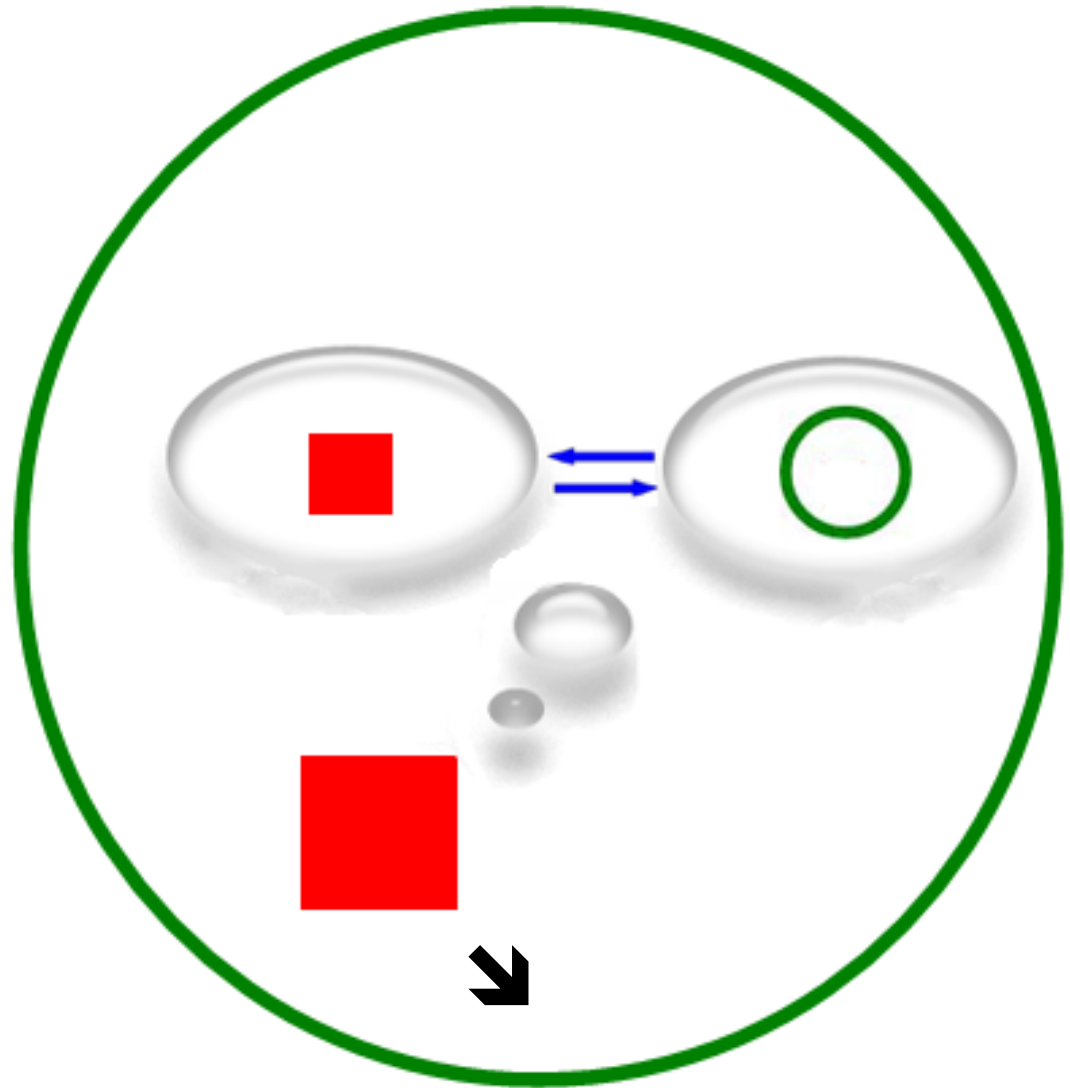


Consciousness and feelings are in the Internal Agent Model (the IAM) – the system's 'software' model of itself...

...and feelings are what influence the evaluative function, enabling the choice of 'good' actions.

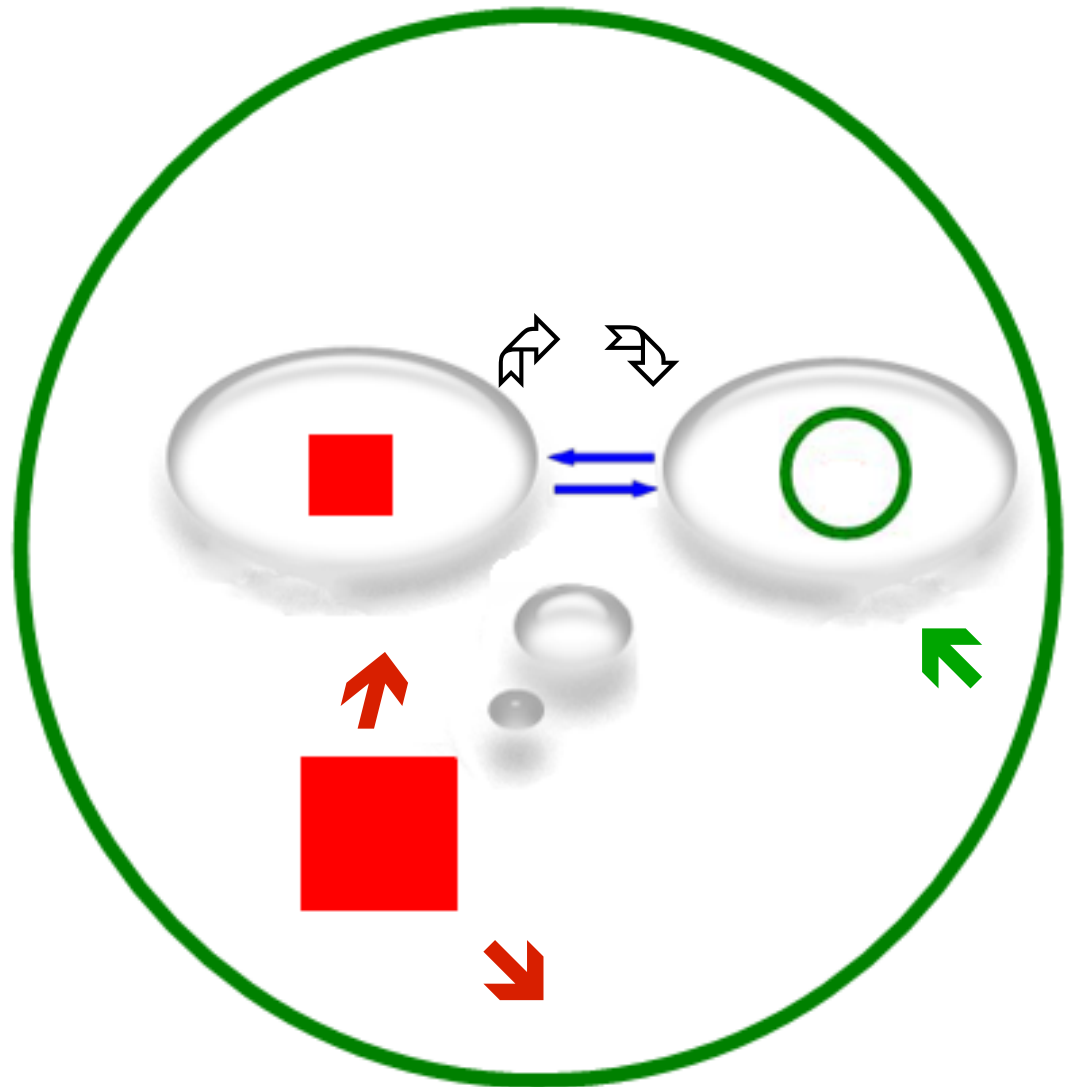


You think you  
control your body,  
and act on the real  
world

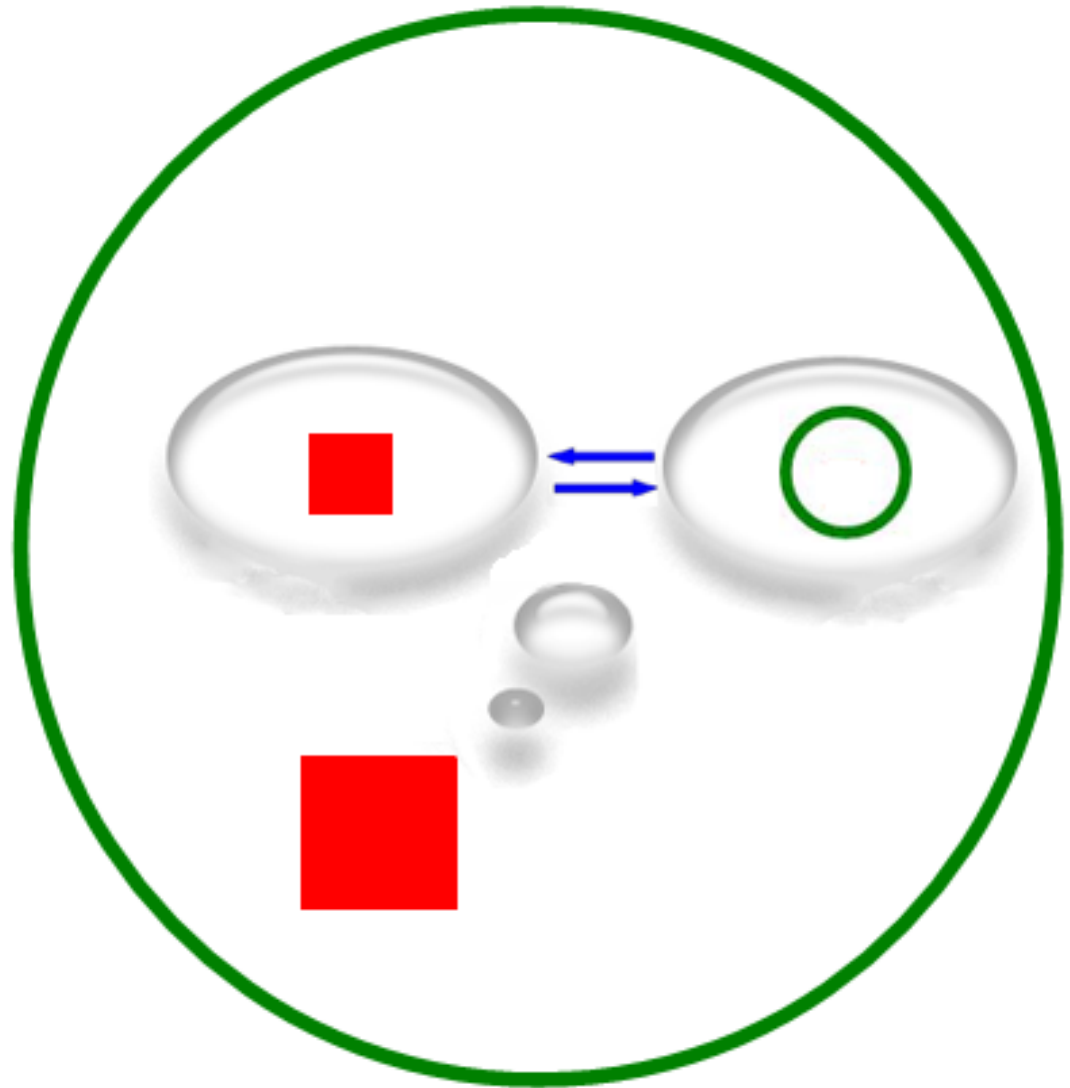


But your body is controlled by other structures within your brain, using the information about 'good' choices.

You attribute its actions to your own agency (or not); this is 'the illusion of conscious will' (Daniel Wegner)



The 'content' of your consciousness is mostly secondary and illusory – it's largely the consequences of keeping the planning system up to date, propagating knowledge through it, and evaluating current and future situations. You occasionally plan, ***but you can never act.***





"The phenomenal self is a virtual agent perceiving virtual objects in a virtual world...I think that 'virtual reality' is the best technological metaphor which is currently available as a source for generating new theoretical intuitions ...heuristically the most interesting concept may be that of 'full immersion'."

Thomas Metzinger 2000

## A proposal

One way to study these phenomena is to build a suitably complex robot, to embed it in a suitably complex environment and to examine the robot's behaviour and internal processes as it learns, evolves, or is designed to cope with its mission.

## A proposal

One way to study these phenomena is to build a suitably complex robot, to embed it in a suitably complex environment and to examine the robot's behaviour and internal processes as it learns, evolves, or is designed to cope with its mission.

And to make sure any internal agent model developed is like our own, we should copy ourselves as best we can – our bodies, as well as our brains.

So how closely should we copy the body?

So how closely should we copy the body?

Sufficiently closely to make it necessary to use motor programs (including those controlling eye movements) qualitatively similar to those used in the human body

So how closely should we copy the body?

Sufficiently closely to make it necessary to use motor programs (including those controlling eye movements) qualitatively similar to those used in the human body

So what are bodies really like?

Not like this!



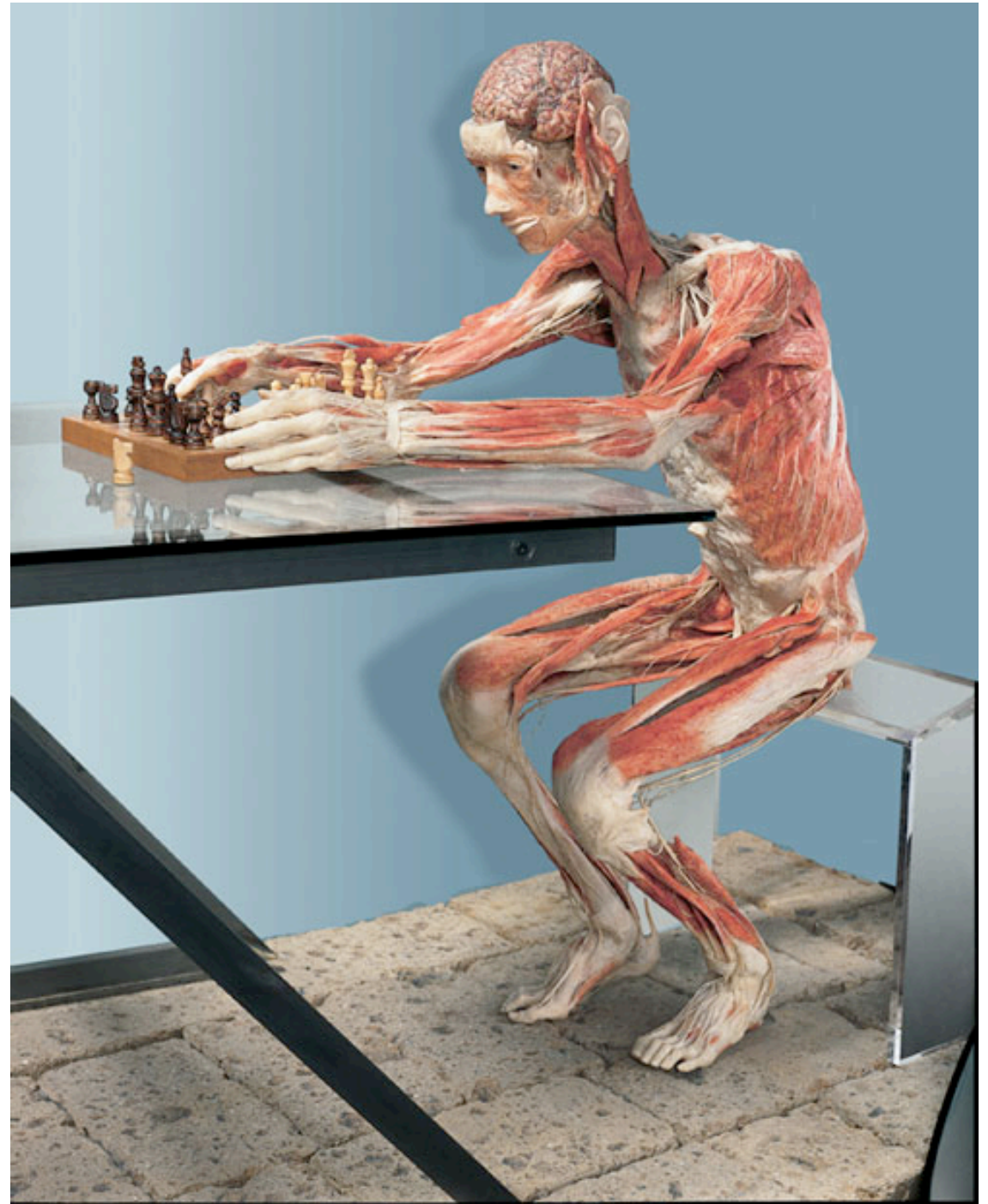
We're familiar enough with the skeleton...





But most people have only become aware of what lies between the skin and the skeleton through the work of Gunther von Hagens.

‘Chess player’



So how closely should we copy the body?

Sufficiently closely to make it necessary to use motor programs (including those controlling eye movements) qualitatively similar to those used in the human body

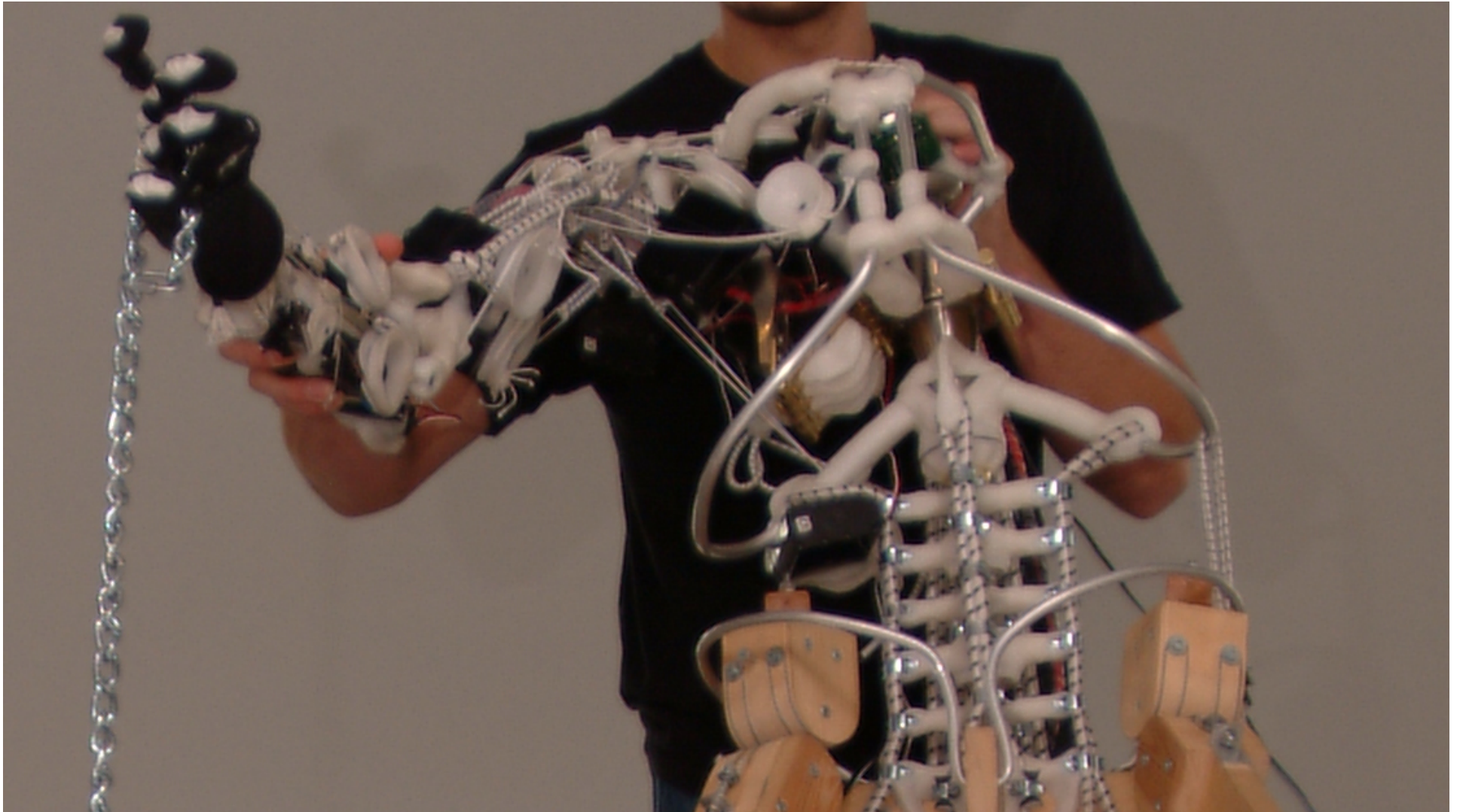
And that means using paired elastic actuators, acting on a body consisting of rigid elements (bones) joined by freely moving joints, and linked by passive elastic elements...

...and you only have to start building robots like that to realise how different they are from 'normal' robots.

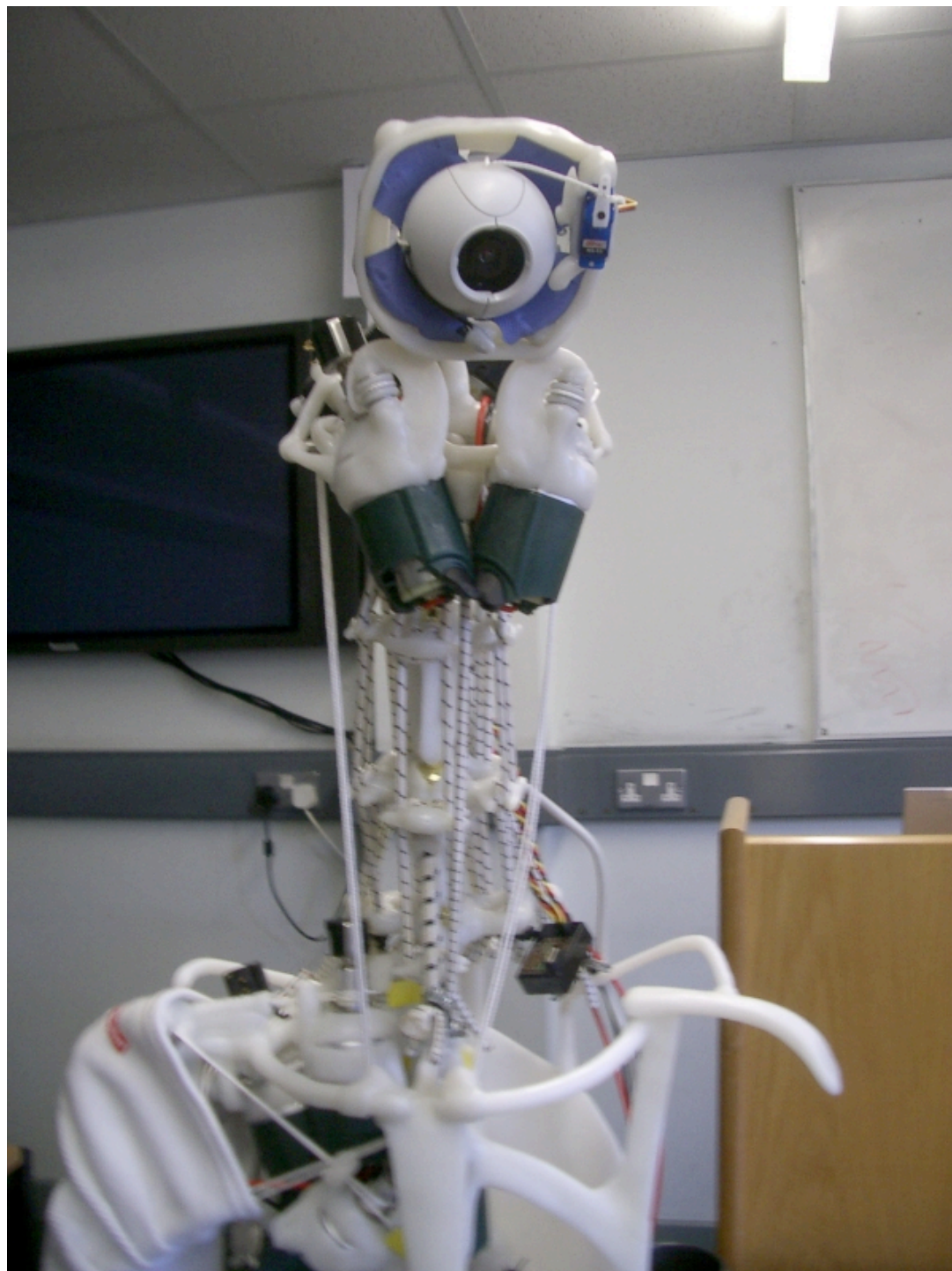
# Copying the body



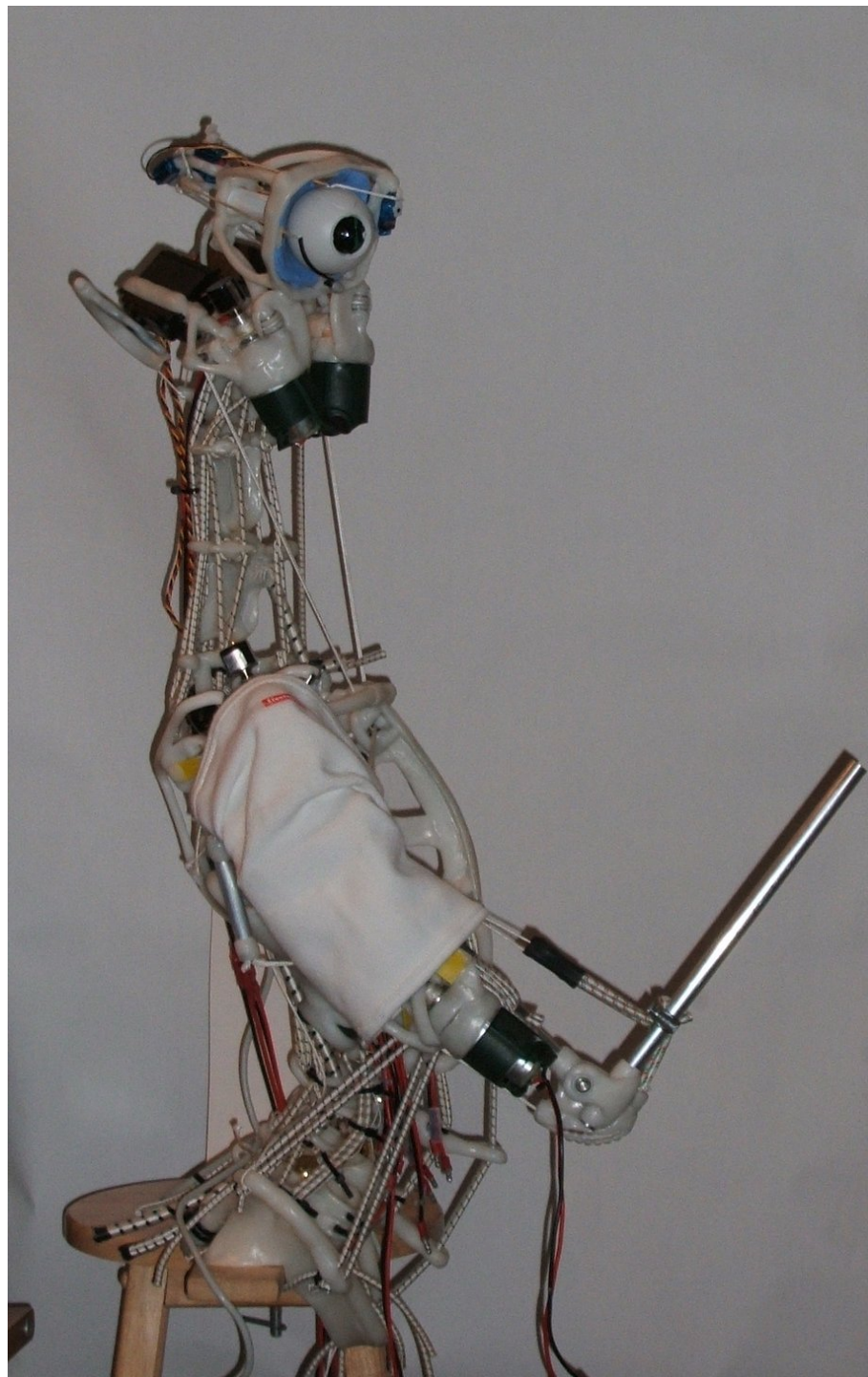
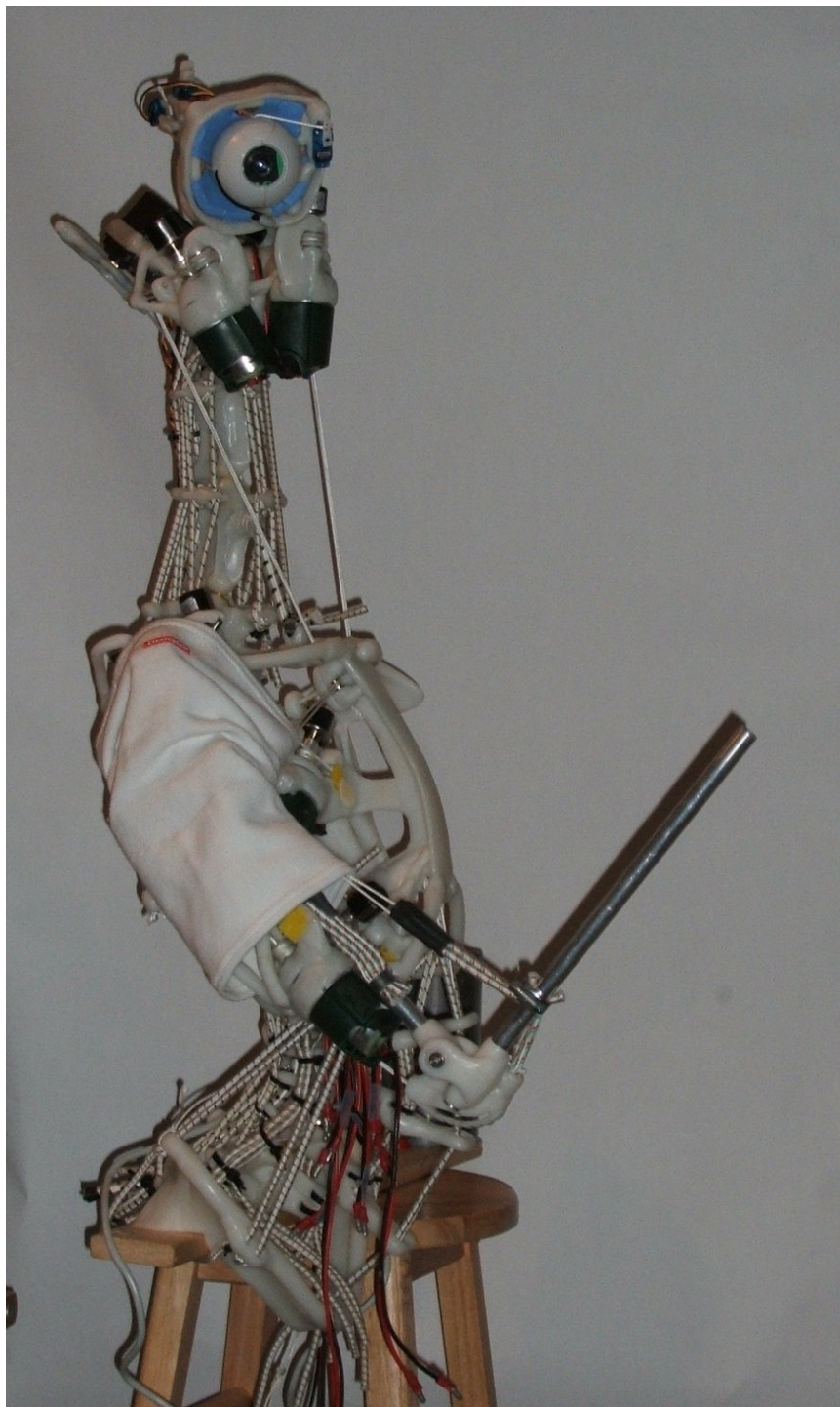




CRONOS 1



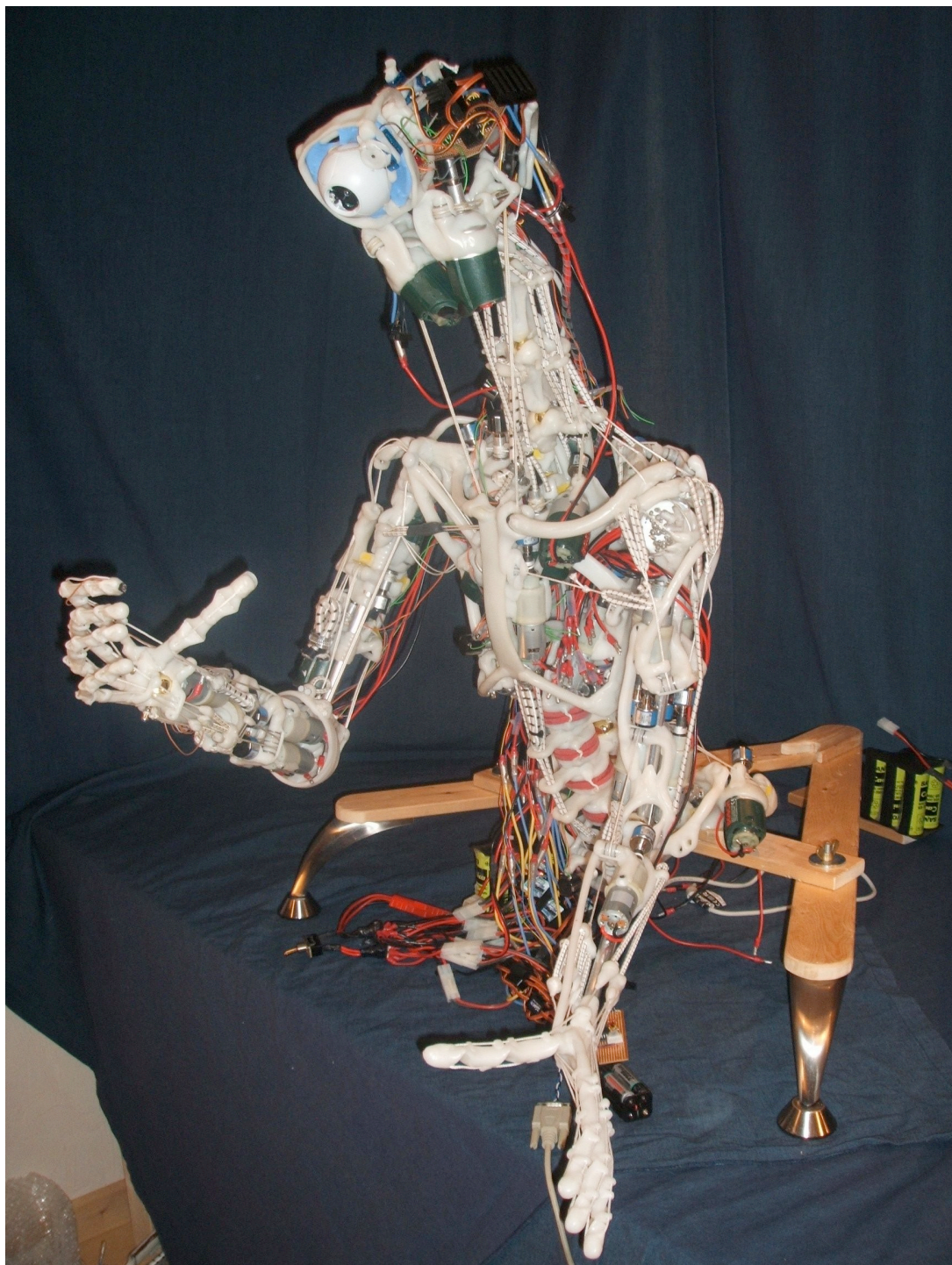




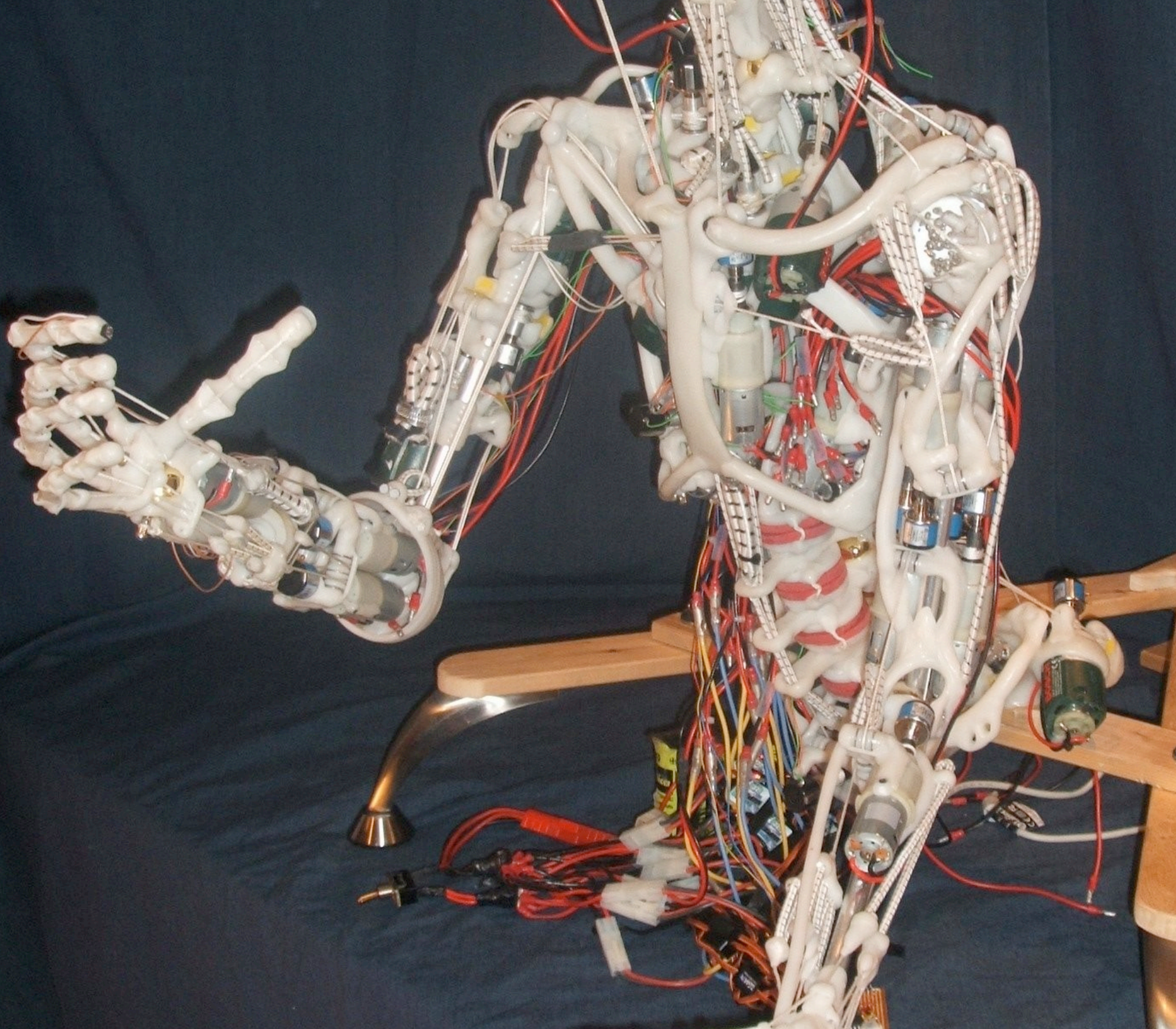




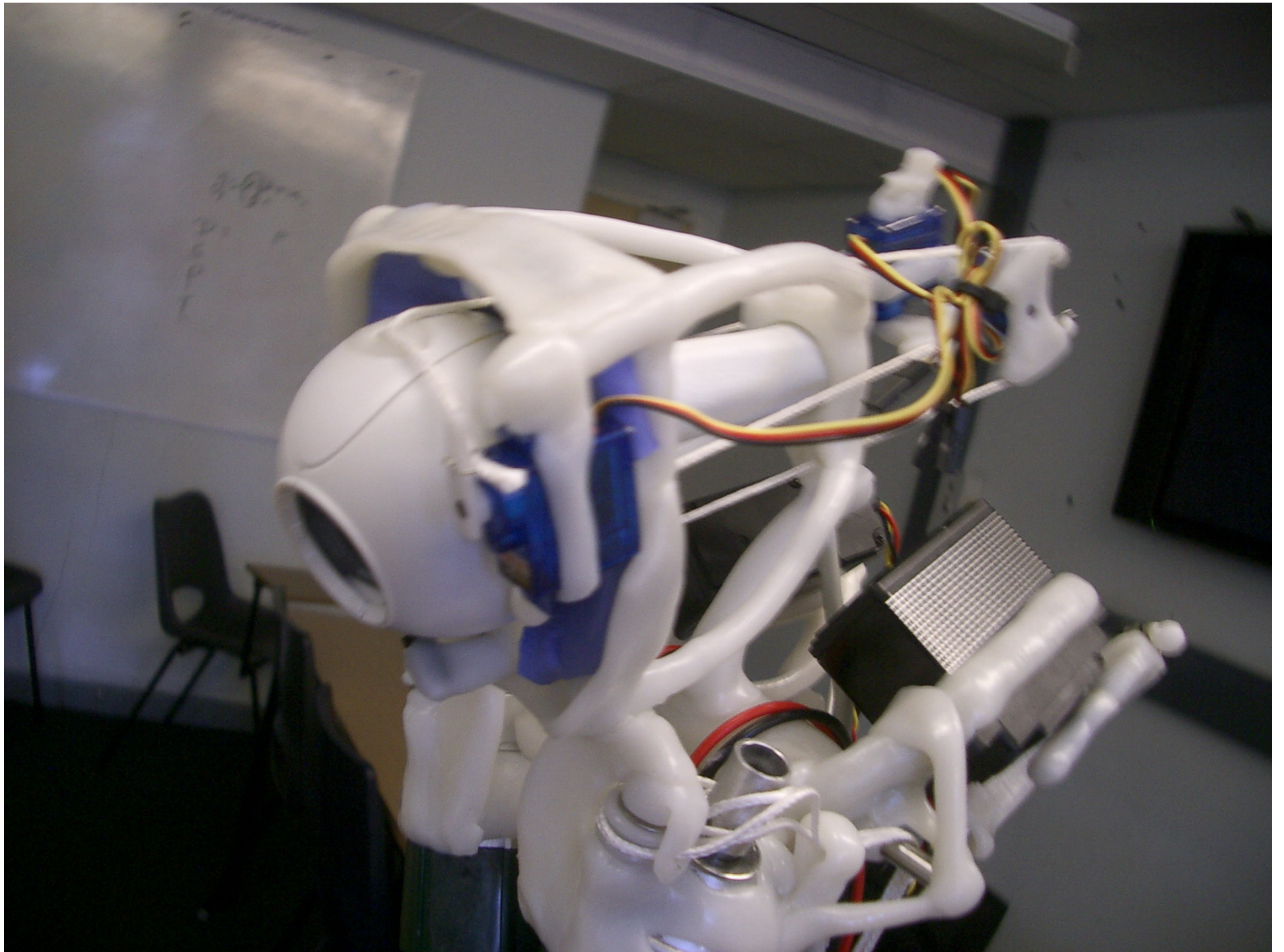




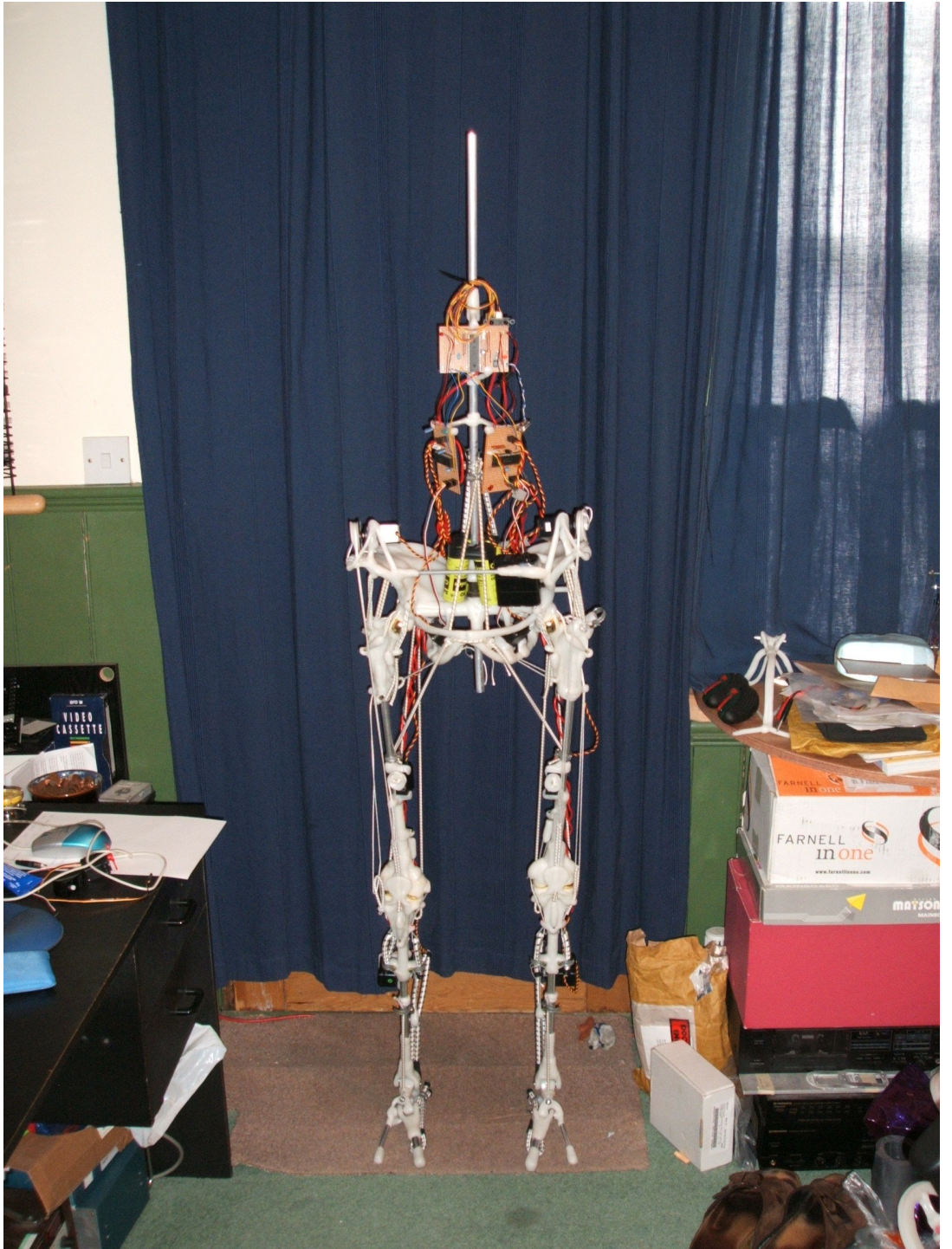












Dem bones, dem bones...

With these ***anthropomimetic*** robots, every movement and every external force is reflected through the whole structure, and they will deform the structure unless active compensation is applied

Dem bones, dem bones...

With these ***anthropomimetic*** robots, every movement and every external force is reflected through the whole structure, and they will deform the structure unless active compensation is applied

*Cue movies*

Dem bones, dem bones...

With these ***anthropomimetic*** robots, every movement and every external force is reflected through the whole structure, and they will deform the structure unless active compensation is applied. Some of this compensation can be reactive, but much of it will have to be ***predictive*** (internal models again!) to enable actions to be carried out from a reasonably stable platform.

Dem bones, dem bones...

With these ***anthropomimetic*** robots, every movement and every external force is reflected through the whole structure, and they will deform the structure unless active compensation is applied

Some of this compensation can be reactive, but much of it will have to be ***predictive*** (internal models again!) to enable actions to be carried out from a reasonably stable platform

This goes far beyond merely maintaining the balance of a passively rigid structure.

## Copying the brain as well

We're also copying parts of the brain – those involved in early vision and the control of eye movements (work being done by Tom Troscianko, Iain Gilchrist, and Ben Vincent, Department of Psychology, University of Bristol)



## Copying the brain as well

We're also copying parts of the brain – those involved in early vision and the control of eye movements (work being done by Tom Troscianko, Iain Gilchrist, and Ben Vincent, Department of Psychology, University of Bristol)

To get started, we're using a saliency mapping system similar to the one that we all have. These systems control what we look at when we're not strongly engaged in a visual task.

## Planning and imagination

Formally, what we are building is a planning system. Previous attempts (GOFAI) used a symbolic substrate, and developed a huge range of methods – more or less mechanical – for generating the actions to be evaluated.

## Planning and imagination

Formally, what we are building is a planning system. Previous attempts (GOFAI) used a symbolic substrate, and developed a huge range of methods – more or less mechanical – for generating the actions to be evaluated.

We know **ALMOST NOTHING** about using a non-symbolic simulation engine as the substrate of a planning system (there is some work using noise to search within learned neural network representations) but we'll take as much as we can from GOFAI.

Is my journey really necessary?

Is my journey really necessary?

If consciousness is about having virtual models of real things, could it also be found in virtual models of virtual things?

Is my journey really necessary?

If consciousness is about having virtual models of real things, could it also be found in virtual models of virtual things?

Couldn't we dispense with the drudgery of building real robots, simulate the whole shebang, and evolve consciousness again?

Is my journey really necessary?

If consciousness is about having virtual models of real things, could it also be found in virtual models of virtual things?

Couldn't we dispense with the drudgery of building real robots, simulate the whole shebang, and evolve consciousness again?

How complex must a conscious system be anyway?

Is my journey really necessary?

If consciousness is about having virtual models of real things, could it also be found in virtual models of virtual things?

Couldn't we dispense with the drudgery of building real robots, simulate the whole shebang, and evolve consciousness again?

How complex must a conscious system be anyway?  
We don't know, but...



## Mammalian brains

Bernard Baars and colleagues have concluded that all mammalian brains contain the neural structures associated with consciousness.

## Mammalian brains

Bernard Baars and colleagues have concluded that all mammalian brains contain the neural structures associated with consciousness.

Human brain: 1400g  $10^{11}$  neurons

Tarsier brain: 4g  $10^9$  neurons

Hog nosed bat brain: 0.07g  $10^7$  neurons



## Mammalian brains

Bernard Baars and colleagues have concluded that all mammalian brains contain the neural structures associated with consciousness.

Human brain: 1400g  $10^{11}$  neurons

Tarsier brain: 4g  $10^9$  neurons

Hog nosed bat brain: 0.07g  $10^7$  neurons

We can do that...

## How will we know if it's conscious?

I don't know, and shouldn't say. But other people are beginning to devise some useful frameworks for answering the question.

Igor Aleksander has proposed 5 axioms to define or characterise consciousness

Thomas Metzinger has identified 11 constraints on "...what makes a neural representation a phenomenal representation"

(T Metzinger, 2003: Being No-one: the self-model theory of subjectivity. 699 pages!)

**AXIOM 1: A SENSE OF PLACE** We feel that we are at the centre of an "out there" world, and we have the ability to place ourselves in the world around us.

**AXIOM 2: IMAGINATION** We can 'see' things that we have experienced in the past, and we can also conjure up things we have never seen. Reading a novel can conjure up mental images of different worlds, for example.

**AXIOM 3: DIRECTED ATTENTION** Our thoughts are not just passive reflections of what is happening in the world - we are able to focus our attention, and we are conscious only of that to which we attend.

**AXIOM 4: PLANNING** We have the ability to carry out "what if" exercises. Scenarios of future events and actions can be mapped out in our minds even if we are just sitting still.

**AXIOM 5: DECISION/EMOTION** Emotions guide us into recognising what is good for us and what is bad for us, and in acting accordingly.

## Metzinger's 11 constraints

- (1) Global availability
- (2) Activation within a window of presence
- (3) Integration into a coherent global state
- (4) Convolved holism
- (5) Dynamicity
- (6) Perspectivalness
- (7) Transparency
- (8) Offline activation
- (9) Representation of intensities
- (10) "Ultrasmoothness": Homogeneity of simple content
- (11) Adaptivity

# What I left out

Architectures

Computational substrate

Social interaction

Language

Creativity



# Conclusions and future work

## Conclusions and future work

Let's face it: there aren't any conclusions yet, and it's almost all future work.

But the important thing is that the work has been started.

## Conclusions and future work

Let's face it: there aren't any conclusions yet, and it's almost all future work.

But the important thing is that the work has been started.

There are now around 50 people actively thinking and working on the issues surrounding machine consciousness. It's almost certain that at least 49 of them are on the wrong track, and I'm likely to be one of their number, but it's the most interesting topic to be wrong about that I can think of.

## Conclusions and future work

Let's face it: there aren't any conclusions yet, and it's almost all future work.

But the important thing is that the work has been started.

There are now around 50 people actively thinking and working on the issues surrounding machine consciousness. It's almost certain that at least 49 of them are on the wrong track, and I'm likely to be one of their number, but it's the most interesting topic to be wrong about that I can think of. **BUT**

## **A warning**

"Suffering starts on the level of Phenomenal Self Models. You cannot consciously suffer without having a globally available self-model. The PSM is the decisive neurocomputational instrument not only in developing a host of new cognitive and social skills but also in forcing any strongly conscious system to functionally and representationally appropriate its own disintegration, its own failures and internal conflicts. Phenomenal appropriation goes along with functional appropriation."

## A warning

“Evolution is not only marvellously efficient but also ruthless and cruel to the individual organism. Pain and any other nonphysical kind of suffering, generally any representational state characterized by a "negative valence" and integrated into the PSM are now phenomenally owned. Now it inevitably, and transparently, is my own suffering. The melodrama, but also the potential tragedy of the ego both start on the level of transparent self-modeling. **Therefore, we should ban all attempts to create (or even risk the creation of) artificial and postbiotic PSMs from serious academic research.**”

T. Metzinger, Being No-One (p 622).

For more information see

[www.machineconsciousness.org](http://www.machineconsciousness.org)

And please feel free to email me

[owen@essex.ac.uk](mailto:owen@essex.ac.uk)

